PhD- FSTC-2017-64
The Faculty of Sciences, Technology and Communication

DISSERTATION

Defence held on 10/10/2017 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN BIOLOGIE

by

Eugen BAUER
Born on 24 May 1989 in Melitopol, (Ukraine)

COMPUTATIONAL ECOSYSTEMS MODELING AND ANALYSIS OF HUMAN-ASSOCIATED INTESTINAL MICROBIAL COMMUNITIES

Dissertation defence committee

Dr. Ines Thiele, dissertation supervisor
Luxembourg Centre for Systems Biomedicine
Associate Professor, Université du Luxembourg

Dr Jörg Stelling
Professor, ETH Zürich, Switzerland

Dr Paul Wilmes, Chairman
Luxembourg Centre for Systems Biomedicine
Associate Professor, Université du Luxembourg

Dr Bas Teusink
Professor, Vrije Universiteit Amsterdam, Netherlands

Dr Ronan M. T. Fleming, Vice Chairman
Luxembourg Centre for Systems Biomedicine
Collaborateur scientifique (Senior), Université du Luxembourg

Molecular Systems Physiology

Luxembourg Centre for Systems Biomedicine

Faculty of Life Sciences, Technology and Communication

Doctoral School in Systems and Molecular Biomedicine

Supported by Fonds National de la Recherche (FNR), Luxembourg (6783162)

Disseration Defence Committee:

Committee members:	A-Prof. Dr. Paul Wilmes
	Dr. Ronan M. T. Fleming
	Prof. Dr. Jörg Stelling
	Prof. Dr. Bas Teusink
Supervisor:	A-Prof. Dr. Ines Thiele

I hereby confirm that the PhD thesis entitled “Computational Ecosystems Modeling and Analysis of Human-associated Intestinal Microbial Communities” has been written independently and without any other sources than cited.

Luxembourg, _____ _____

Eugen Bauer

"All models are wrong but some are useful."

George Box

Acknowledgments

I first want to thank all that helped me during the projects of my thesis. I thank the Fonds National de la Recherche for giving me the opportunity for my PhD by their generous funding. My supervisor, Prof. Ines Thiele, gave me enough freedom to develop my ideas during my PhD and I enjoyed the scientific discussions with her. Dr. Cedric Laczny, Dr. Stefania Magnusdottir, and Prof. Paul Wilmes helped me during my first project by giving new ideas and additional directions. I also want to thank Paul for his comments during the numerous meetings during my PhD. Johannes Zimmermann, Prof. Christoph Kaleta, and Federico Baldini were instrumental for the development of the software package "BacArena". In particular, Johannes co-developed the package with me and I really enjoyed the philosophical discussions with him about our key modeling concepts and assumptions.

Next, I want to thank all my colleagues at the Luxembourg Centre for Systems Biomedicine for giving me a nice working atmosphere. I also want to thank the Molecular Systems Physiology group for hosting me during my PhD. Especially, I acknowledge Federico Baldini, Alberto Noronha, and Marouen Ben Guebila for the scientific discussion, but I also want to thank them for their personal support and friendship. I want to extend this acknowledgement to the Regional Student Group Luxembourg, of which I could proudly serve as a president in the last year of my PhD. I thank the board of this student group for their motivation and creativity during the organization of our science and networking events, which were always a lot of fun for me.

Finally, I want to thank my parents Wadim and Larissa Bauer for their trust in me and their continuous support, without which I would not have been able to complete my studies. I also want to thank my friends Artur Kunz and Ana Correia for their support and advice during my PhD.

Contents

List of Figures

List of Tables

List of Abbreviations

ABM	Agent based modeling
AGORA	Assembly of gut organisms through reconstruction and analysis
ATP	Adenine triphosphate
CO2	Carbon dioxide
COBRA	Constraint-based reconstruction and analysis
COG	clusters of orthologous groups
CD	Crohn's disease
dFBA	Dynamic flux balance analysis
DNA	Deoxyribonucleic acid
EC	Enzyme commission
FBA	Flux balance analysis
FNR	Fonds national de la recherche
FVA	Flux variability analysis
gDW	Grams dry weight
GPR	Gene-protein-reaction
IBD	Inflammatory bowel disease
IBM	Individual based modeling
PCoA	Principle coordinate analysis
RMSE	Root-mean-square error
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
SBML	Systems biology markup language
SCFA	Short chain fatty acid
SIHUMI	Simplified human intestinal microbiota
t-SNE	t-distributed stochastic neighbor embedding
VMH	Virtual metabolic human

Summary

The human microbiota consists of several hundred bacterial taxa that engage in complex ecological interactions with each other by exchanging metabolites. A loss of taxonomic and metabolic diversity of this system is often associated with various diseases such as inflammatory bowel disease. Approaches with omic technologies have been used to elucidate the taxonomic and functional components of the human microbiota. However, it is still challenging to link the different components with each other to understand the functionality of each bacterium in its ecological context. Therefore, various systems biology approaches emerged which investigate the metabolic interactions within microbes of a community.

The goal of this PhD thesis is to develop and apply methods in systems biology to model the metabolic complexity of the human gut microbiota. In a first study, the individual metabolism of several hundred bacteria has been automatically reconstructed and analyzed to investigate relevant metabolic differences between species. The second study focused on building a bottom up framework that combines constrained based and agent based modeling to simulate the metabolism of individual bacterial cells that can interact in terms of different species populations. This framework has been calibrated and applied to analyze a simplified human microbiota consisting of seven species. In a final study, the developed framework was further integrated with metagenomic data of the gut microbiota from patients with inflammatory bowel disease as well as healthy controls to model disease associated differences and predict novel treatments.

Chapter 1

Introduction: Network Representation and Modeling of the Human Gut Microbiota

1.1 Introduction

Microbial communities are abundantly present throughout nature and form many beneficial symbiotic interactions with different eucaryotic hosts. In many cases, the host acquires novel functionalities with symbionts that can contribute to host fitness and wellbeing [43] by, e.g., the supplementation of essential nutrients [165]. These functional microbes can be often found on the interface of nutrient absorption in the intestinal ecosystem [40] and gut associated structures, such as bacteriomes [20]. Removal of these symbiotic microbes can lead to host fitness decrease, demonstrating the host dependence on the symbionts [165] and the relevance of symbiotic interactions.

While the gut ecosystem of human hosts is less well understood than other symbioses, intestinal microbial communities are being associated with human well-being [30]. In particular, microbiota metabolism is considered relevant for nutrient provisioning and complementary digestion of food [39]. The human gut microbiota consists of more than thousands microbial species [222], which can form manifold metabolic interactions between themselves and with their host. Interestingly, the microbiota metabolism is more conserved between

human individuals than the species composition, suggesting redundant functionalities are present between microbes and complement each other [88]. The metabolic functions of various gut microbes are suggested to benefit human health with the provisioning of vitamins [121] and fermentation products [151]. Microbial fermentation products can be utilized by the human host as an additional energy source [38] and benefit the immune system [59], therefore playing a pivotal role for human well-being.

A loss of microbial and metabolic diversity can lead to various microbiota associated diseases, such as obesity [218], type 2 diabetes [159], and inflammatory bowel disease (IBD) [58]. IBD, for example, is characterized by an inflammation of the gut epithelium and a dysbiotic gut microbiota, which is decreased in species richness and metabolic functions [158]. This has a profound effect on the concentration of fermentation products, which are decreased in IBD patients compared to healthy controls [86]. Short chain fatty acids, in particular, are thought to influence IBD symptoms [164]. These metabolites are subject to cross-feeding interactions between a variety of microbes [12], which further underlines the relevance of a rich microbiota composition with different metabolic functionalities.

Treatments for gut associated diseases usually try to modulate the human gut microbiota to exhibit healthy characteristics. Such treatments can include fecal microbiota transplantation, probiotics [60], and change in diet in the form of prebiotics [22]. Since the microbiota varies between individuals, treatments have to be personalized to support each patient individually. Furthermore, it is important to understand the mechanism by which the microbiota is influenced to design novel treatment with higher efficacies.

With the advent of high throughput sequencing technologies, it is possible to enumerate and describe the microbial and functional diversity of the human gut with respect to different diseases and conditions. These analyses unraveled a complex gut microbial ecosystem that is influenced by diet and environmental factors [36]. Furthermore, transcriptomic and proteomic analyses allowed to probe the activity of the microbiota, revealing high activity of fermentative pathways and carbohydrate utilization [223]. Whole genome sequencing of single microbes can give hints on the functions and capabilities of individual members of the intestinal microbial community [205]. Recently, it became also possible to reconstruct individual genomes from metagenomic data [166]. Taken together, omics analyses have broadened our understanding of the human gut microbiota in terms of the possible metabolic

function and microbes that occur in this complex ecosystem. However, it is still difficult to assess the ecology in terms of metabolic interactions between microbes and how each microbe contributes to the intestinal microbial community. To unravel the ecological mechanisms that drive the human gut microbiota, it is important to go beyond the descriptive nature of high throughput data analysis. This can be achieved by formulating knowledge or data driven models, which can describe the underlying ecosystem and give mechanistic insights for hypothesis generation and designing further experiments. A recent review described the relevance of such models in giving novel causal relationships in a field that is flooded by data and correlation [211].

Here, we will discuss current systems biology approaches that have been applied to the human gut microbiota. As highlighted above, metabolism is one of the key features in the gut microbiota and microbes in general, we will thus focus on this aspect. First, we outline network topology analyses that are based on newly generated data, or already existing knowledge. Then, we will describe modeling approaches in the field of constraint based modeling that allow to simulate interactions within microbial communities. Finally, we will highlight the advantages as well as limitations of these approaches and give suggestions for further studies.

1.2 Network based approaches applied to the human gut microbiota

Network based approaches are used to identify relevant microbes or metabolites of the human gut microbiota. Networks are usually represented by interactions in form of edges that connect biological components in form of nodes. With respect to the human gut microbiota, these components can represent species that interact [52], or metabolites that are converted through reactions [183]. The goal of such networks is to provide a global overview of the underlying system and possible mechanisms, which make it possible to identify relevant microbes or metabolites that play important roles in the network by connecting a variety of components or exhibiting key features. In the next paragraphs, we will discuss networks that are created top-down with newly generated data and knowledge derived networks that are created buttom-up

with existing information (Figure 1.1).

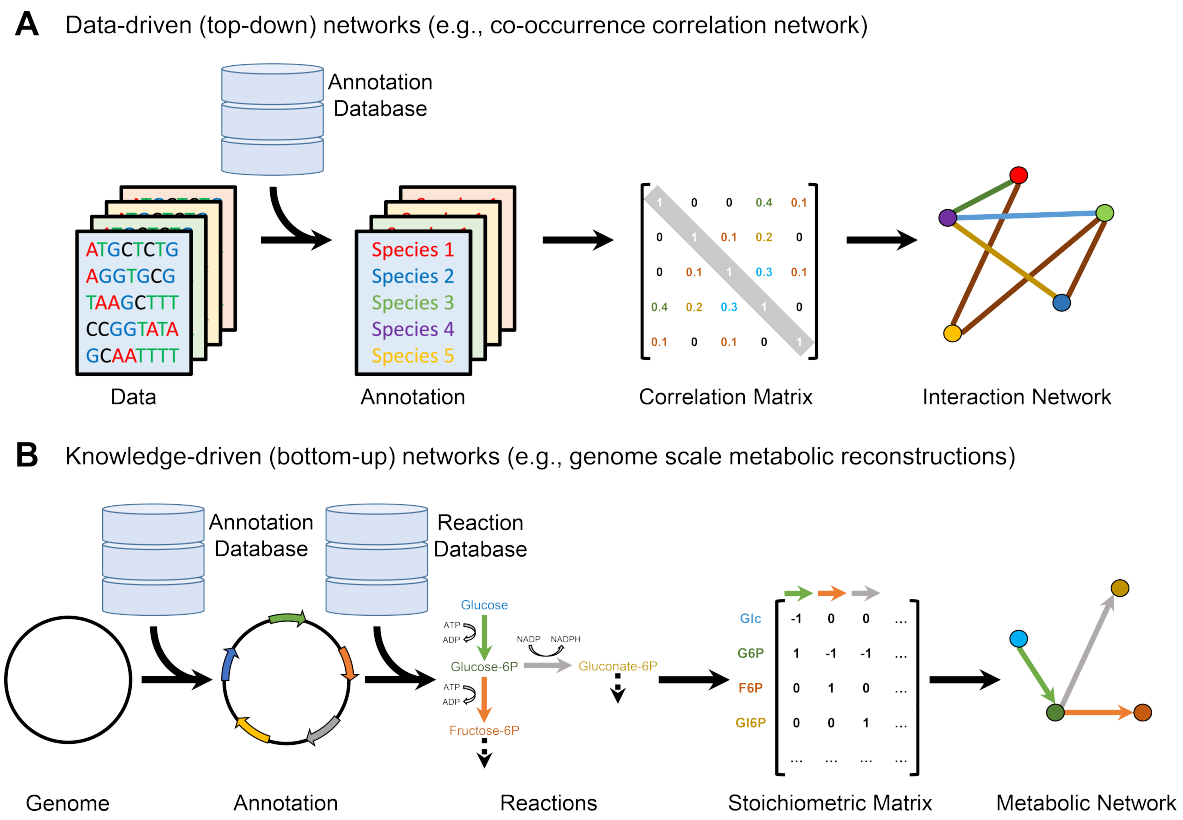

Figure 1.1: Examples of data-driven network reconstruction based on high throughput data from different patients or conditions (A) and knowledge-driven networks based on genome scale metabolic reconstructions (B).

1.2.1 Data-driven (top-down) networks

Species co-occurrence networks can be used to investigate the ecological interactions between microbes. Based on high throughput data analyses, simple networks can be constructed with the species abundance information for different patients [52]. Such networks contain the information of which species are co-occurring by calculating correlation coefficients for each pair of microbes (Figure 1.1A). Consequently, a positive interaction can be deduced if two microbes have a positive correlation, e.g., co-occur with each other, and a negative interactions if the coefficient is negative, e.g., exclude the presence of each other [52]. This information can be relevant for understanding the ecological concepts that drive exclusion mechanisms and niche differentiation. Furthermore, the information on what species interact positively or negatively can drive further experimental studies that investigate the mechanism by which

these interactions take place, e.g., production of antibiotics or metabolic cross-feeding. This also highlights the limitation of such network based approaches, since they do not provide the hypothesis for the mechanism of species interactions.

Metabolism based networks can be useful to assess the functional aspect of the human gut microbiota. Metabolic functions or enzymes annotated from high throughput sequencing of different human individuals can be used to find pathway or metabolite differences [72]. By connecting enzymes based on the reactions that they are involved in, it is possible to construct metabolic networks specific for each patient [72]. Topological analyses of these networks can reveal patient specific differences as well as global differences between healthy controls and patients [72]. Additional information can be gained by finding differences in enzymatic modules consisting of multiple related metabolic functions, which are connected and thus influence each other. This can be important when trying to find potential components for treatments, or assessing off-target effects [72]. While such networks are primarily constructed from patient data, the connections between the reactions are retrieved from previous knowledge, e.g., databases (Figure 1.1).

1.2.2 Knowledge-driven (buttom-up) networks

Metabolic networks that are primarily constructed from previous knowledge can be used to generate an overview of metabolic functions. Such networks rely on databases that store biological information. For metabolic networks, the most commonly used databases are KEGG [95], MetaCyc [27], and ModelSEED [84]. These databases store biochemical and sequence information about metabolic reactions and their related enzymes. This information can be used to construct a global view of the metabolic reactions and pathways that take place in the human microbiota to assess its overall metabolic capacity [223]. Useful visualizations of these networks can be achieved by maps representing how reactions interact with each other and share metabolites [223]. By sub-selecting these maps according to patient data, it is possible to find differences in terms of various gut locations [223]. With these maps it is possible to link metabolic functions and generate an global overview, however, the information on which metabolic functions are carried out by specific microbes, and how they interact is missing in such analyses.

In a recent approach, the metabolic exchange between microbes was represented in a global interaction map which can help to identify deficiencies in diseases [187]. Based on databases and scientific literature, microbes are essentially represented by their transport of various metabolites which can act beneficial or detrimental to fellow species. The global view of all transporters gives some clues on how nutrients are converted by specific microbes and then supplemented to the host. When mapping patient data to this network of transporters it is possible to identify potential exchange deficiencies compared to healthy controls [187]. While these analyses can reveal potential metabolites or interactions that are differentially regulated in patients, they lack the view on the complete metabolism of each microbe species.

Genome scale metabolic reconstructions can be used to represent the metabolism of an organism to investigate the metabolic potential of gut microbes. Automatic pipelines such as ModelSEED [84] can generate such reconstructions by the genome annotation, which can be retrieved from RAST [147]. Based on the annotation of each gene, enzymes are predicted, which carry out one or multiple reactions and vice versa. The reactions are then retrieved from a database such as KEGG [95], or MetaCyc [27], and are represented in a stoichiometrically accurate manner by their metabolite educts and products as well as their thermodynamic directionality (reversible or irreversible). The complete set of reactions that are retrieved by this process constitute the metabolic reconstruction and are mathematically represented in form of the stoichiometric matrix (S-matrix). Rows in this matrix represent metabolites and columns reactions. Entries are stoichiometric coefficients for each metabolite of each reaction. Through sharing different metabolites, reactions are connected with each other and therefore represent a metabolic network. In Chapter 2, we applied the automatic pipeline of ModelSEED [84] to reconstruct 300 representative microbes that are present in the human gut [9]. We compared the metabolic networks with each other to find taxa specific differences between the gut microbes. Based on this analysis we found that microbial strains can be more metabolically different than predicted by phylogeny, which highlights the need for taking a diversity of microbes into consideration to understand the complete metabolism of the human gut microbiota.

1.3 Constraint based modeling of intestinal microbial communities

1.3.1 Constraint based reconstruction and analysis (COBRA)

Constraint based reconstruction and analysis (COBRA) is based on genome scale metabolic reconstructions, which are formalized as metabolic models to simulate biologically relevant physiological states. By applying condition specific constraints as well as a metabolic objective to the S-matrix, reconstructions are converted into metabolic models (Figure 1.2A). This information is then formulated as a mathematical problem that is used to retrieve the vector of reaction fluxes v by

$$\begin{gathered} \text{maximize } v_B \\ \text{subject to: } S \times v = 0 \\ v_{i,min} \leq v_i \leq v_{i,max} \end{gathered}$$

for all reactions i and maximization of biomass flux v_B under the steady state assumption, where metabolites cannot accumulate and the total flux into the network must equal the total outflux. The constraints can represent medium conditions in which the uptake of metabolites via exchange reactions is limited or limitations of internal reaction fluxes that come from experimental data [192]. The biomass flux then predicts how much biomass the organism can produce under the given conditions. Fluxes that flow through the network predict the metabolic pathways that are used by organism in order to achieve the objective, e.g. biomass production. The process of finding the solution to the stated optimization problem is called flux balance analysis (FBA) and can be solved via linear programing. To help with the interpretation and data integration, fluxes are scaled to a unit of mmol per gram dry weight per hour and the biomass is usually given in gram dry weight per hour [145]. Constraint based modeling approaches can be applied to the human gut microbiota by either analyzing the metabolic capabilities of single species [9] or combining metabolic models of multiple species in a community modeling approach [14].

COBRA approaches can assess the spectrum of metabolism for an organism. Since the system of linear equations for FBA calculations on most metabolic models is usually under-determined, e.g. more reactions/variables than metabolites/equations, multiple flux

distributions can achieve the given solution. To subselect the solution space, parsimonious FBA can be used to select the solution that minimizes the total reactions flux, which is an estimate for minimization of enzyme usage [111]. This can be computed by a separate linear programming problem, which minimizes the total flux while ensuring the calculated biomass objective [111]. Furthermore, flux variability analysis (FVA) can be applied in which each reaction is each minimized and maximized in an iterative manner to find flux spans that can be carried by each reaction [122]. These calculations enable the assessment of metabolic functionalities an organism can achieve.

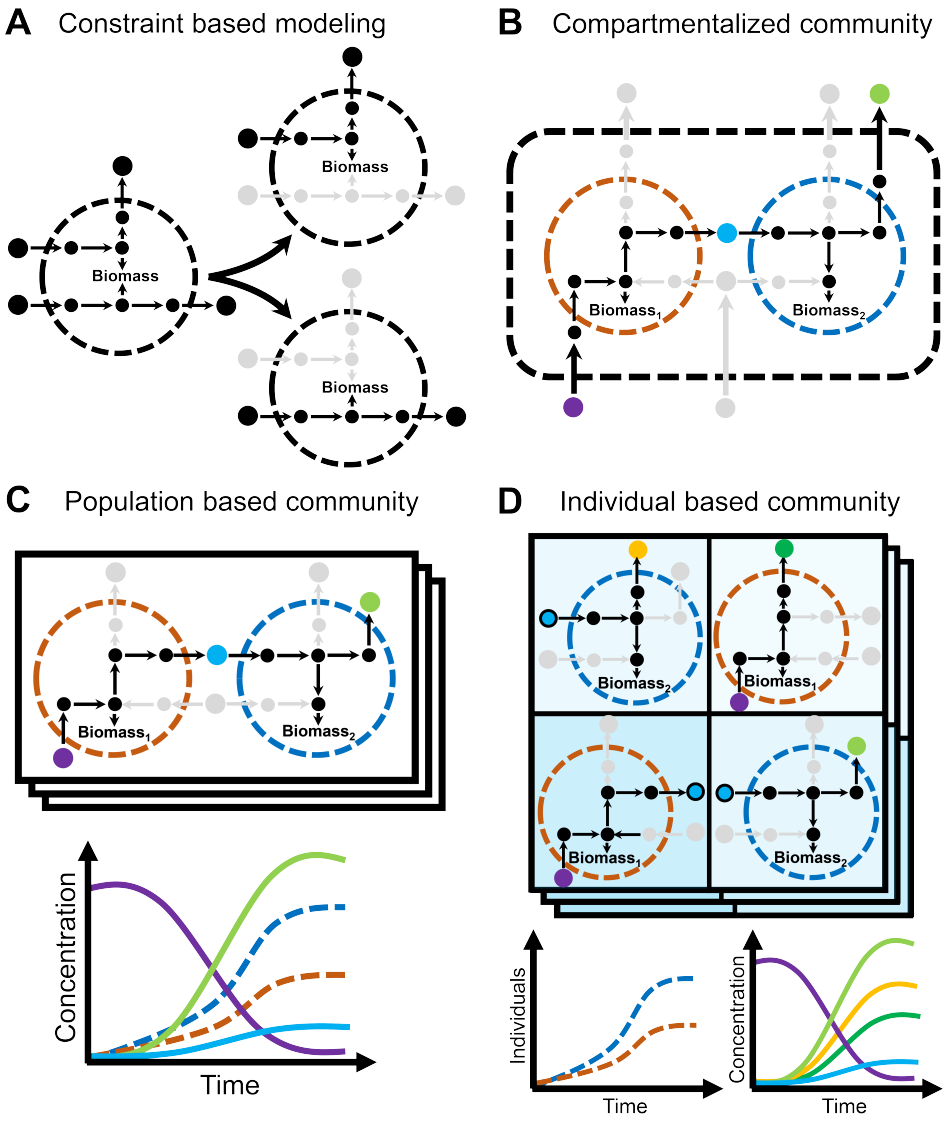

Figure 1.2: Flux balance simulation (FBA) of constraint based models (A) and compartment based communities (B) as well as dynamic FBA of population based (C) and individual based models (D).

1.3.2 Compartmentalized community models

Static community models can be simulated by integrating microbe reconstructions with each other via their S-matrices (Figure 1.2B). In this combined model, the individual microbes are separated from each other by occupying different compartments, in which they can secrete and take up metabolites in a shared environment [101]. Therefore, the microbes can compete for nutrients in this environment, but can also support each other by releasing metabolites that can be used by the other microbe. The optimized biomass of this community is usually composed as a combination of the individual biomass reaction of each microbe [101]. Additionally, coupling constraints can be applied to ensure that the biomass of each microbe is scaled by its uptake of nutrients [80]. Further developments of the community objective include a multi-objective optimization in which the egoistic growth interest of each microbe is optimized while ensuring an overall community growth [225]. These methods allow the prediction of metabolic interactions based on combined S-matrices.

Pairwise models of microbial communities can give new insights in terms of the metabolic interactions that occur in the human gut. In a recent study, we applied pairwise community models to the human gut microbiota by modeling all possible pairs of over 700 microbes and found strikingly mostly negative metabolic interactions, e.g. competition for resources, of different microbes [183], which underlines the ecological concept that negative interactions shape diversity in the human gut microbiota [34]. By simply adding additional microbes to the compartment based strategy, it is also possible to model intestinal microbe communities of multiple species [82]. A study based on 11 metabolic models [81] revealed a vivid exchange of fermentation products and the conversion of metabolites involved in neurotransmitter production, which can have important implications in various neurological diseases. While such generated information of compartmentalized community models can be useful in finding new hypothesis of metabolic interactions, the predictions rely on several assumptions such as a joint community growth and steady in- and outflux of metabolites.

1.3.3 Population based model dynamics

Population based COBRA modeling can be used to analyze temporal dynamics of metabolic interactions within microbial communities. With the optimization of biomass growth FBA

intrinsically models growth of microbe populations. The combined S-matrix approach can be also used to model population dynamics with respect to a temporal scale [221]. Essentially, the biomass optimization of each microbe is solved iteratively and independently. Each iteration represents a discrete time interval in which the biomass is produced and metabolites are secreted or taken up. The units of the fluxes and biomass are scaled accordingly to this time interval. Similar to the compartmentalized community models mentioned above, metabolites are secreted into a shared compartment. Therefore, the same positive or negative interactions between microbes as discussed above can potentially take place. Furthermore, through the integration of a temporal scale via time intervals, it is possible to investigate the evolution and emergence of these interactions under different conditions. This can help with the understanding of how different interactions are temporally dependent, as it is thought to be the case for the human gut microbiota where nutrient intake is dynamically changing throughout the day [200]. Experimental time series data can be integrated in such temporal dynamics to adjust the simulation results [118]. Recent approaches also add a spatial dimension to such community models to represent the colony growth of organisms [79]. The inclusion of a spatial dimension allows the representation of metabolite concentrations which can diffuse and create different gradients. Such gradients are thought to strongly influence the gut microbiota by forming different niches, in which the microbial community differentiates [46]. Population based community modeling can therefore help to investigate what metabolites affect and dynamically change the microbial community.

Population based COBRA modeling can give new insights in the dynamic change of metabolic interactions of human gut microbial communities. A recent approach applied population based metabolic models to a simple community of six microbes [71]. The results of this analysis demonstrated a highly metabolic active community that exchanges a variety of different metabolites through a complex interaction network. Within this interaction network many cross-feeding interactions have been validated against knowledge from literature. Notably, these interactions can change over time and are therefore highly dynamic [71]. A significant assumption of such models is that each microbe consist of a population of homogenous cells which operate under similar conditions. This can be useful to investigate the metabolic behaviour of populations, but is limited in predicting metabolic interactions between individuals of a population.

1.3.4 Individual based community models

The metabolic interactions between individual organisms can be modeled by combining COBRA with individual based modeling (IBM). IBM is a method also used in classical ecology [91] to model species populations in terms of single independent individuals in discrete time steps and a spatial environment. Individuals in this environment can interact according to predefined rules, which determine their states. Through these local interactions, global population structures can emerge that determine the system state. Therefore, population dynamics can be explained by the individuals by which they are made up, which helps in understanding how single cells shape and structure populations of species in a ecosystem. Several approaches combine IBM with constraint based modeling to study the metabolism of single cells in a population of single [15] or multiple species [10, 204]. Essentially, each single cell is represented by a metabolic model that simulates its metabolism according to the spatial position in the environment, which includes metabolites. Similar to population based models described above, metabolites can diffuse through this environment, e.g., by partial differential equations, to create different concentration gradients that dominate the community by creating niches in which different metabolic pathways are activated. Each species is thus represented by metabolically heterogeneous individuals that could simulate the full metabolic potential of that microbe [10]. This allows to predict the metabolic interactions between species as well as within species, which can have different metabolic phenotypes depending on the spatial resource allocation (Figure 1.2D).

Individual based COBRA modeling can give insights in the spatial and temporal community structure of the human gut microbiota. In Chapter 3, we developed and applied this combined approach of IBM and COBRA to analyze microbial communities in the human gut [10]. Our community was simplified to seven representative microbes that were also previously experimentally tested [11]. With our simulations, we could recapitulate experimentally known metabolite concentration, but could predict also novel cross-feeding interactions with fermentation products that were vividly exchanged within the microbes of the community [10]. Furthermore, by applying a spatial gradient of mucus glycans we could observe a spatial niche differentiation of microbial cells as would have been expected from experimental microscopy studies. This further strengthens the fact that metabolism is an important factor in

shaping the gut microbiota and inducing ecological interactions. Such findings demonstrate the relevance of integrating ecological methods with COBRA to understand the metabolic mechanisms that shape the temporal and spatial community structure.

1.4 Current challenges of gut microbiota modeling

COBRA approaches for analyzing microbial communities are promising tools for investigating the metabolism of the human gut microbiota. Several studies have been conducted in this context, mostly focusing on small communities representative for the intestinal microbiota (Table 1.1). These studies revealed several important aspects of the human gut microbiota and its metabolism by providing mechanistic models, which allow to trace metabolic interactions between species. These analysis can improve our understanding of the metabolic mechanisms that shape intestinal microbial communities, but there are several limitations and challenges that need to be considered when applying COBRA community modeling to the human gut microbiota.

Table 1.1: List of the different constraint based community modeling approaches that have been applied to model microbial consortia of the human gut microbiota.

Strategy	Application to the human gut microbiota	Species	Ref.
Compartment based models			
	Host-microbe metabolic interactions	2	[80]
	Human metabolic interactions with microbial community	12	[82]
	Metabolic interactions between microbes in community	11	[81]
Population based models			
	Dynamic metabolic interactions within microbial community	6	[71]
Individual based models			
	Niche differentiation induced by mucus glycans	7	[10]
	Diet interactions with microbiota and metabolic cross-feeding	3	[204]

1.4.1 Scalability and model complexity

One of the most striking hallmarks of the human gut microbiota is its species diversity, which poses challenges to model simulations. Simplified microbiota models of <10 species are relevant for studying metabolic interactions in general and can be used to simulate

experiments that are conducted with small microbial communities in gnotobiotic animals [11] or in vitro [173]. However, such models will never be able to capture and explain the complexity of the human gut. Why are there so many different species? Why is the human gut microbiota more diverse than other body sites? Why are some diseases associated with a lower microbiota diversity? These are some of the questions that can be only addressed by a more comprehensive microbiota model. In a recent publication [183], we created a resource of over 700 curated metabolic models of gut microbes. Combining those into a community model poses several difficulties such as the time of simulating. For addressing the simulation time, the community models need to be scalable, e.g. the simulation time should be independent from the number of species. Compartmentalized models are extended versions of single metabolic models, which become more difficult to solve since the number of variables is increasing. In particular, extensive calculations such as FVA require more sophisticated simulation algorithms that allow for instance parallelization [83]. Population based models simulate each microbe independently and scale therefore with the number of species in the community. Individual based community models are generally independent on the number of species but scale with the number of individuals linearly [10] and are therefore limited in modeling a small spatial scale. Taken together, each simulation experiment must be performed with considering a tradeoff between model complexity and simulation time depending on the modeling paradigm.

Model complexity also poses the problem of data analysis. By simulating large scale microbiota models with a high number of variables, it becomes difficult to explain mechanistically how these different parameters and changing conditions influence the overall results. It is therefore important to simulate various conditions to account for the stability of the system. The large amount of simulations and the complexity of the models further require data mining approaches to find the most relevant parameters influencing the system. This reduces the complexity of the simulations to more simple hypothesis that can be tested in further analysis and with experiments.

1.4.2 Data integration

Data from omics experiments can be integrated into COBRA community microbiota models to generate context specific models. Given the avalanche of high throughput data from human gut microbiota studies, it becomes more and more attractive to integrate these data in community models. In our recently published resource of gut microbe reconstructions [183], we mapped metagenomic data of microbial abundances onto our set of microbes, which resulted in microbiota sizes of about 100 microbes. Metagenomic data information of microbe species can be integrated by scaling the biomass of the corresponding microbe in the community model to the calculated abundance. Moreover, samples of different patients or conditions can be integrated in such a model to simulate condition specific microbiotas. If available, further data such as transcriptomic or metabolomic information could be integrated to modulate the activity of microbes. By their genome-scale representation of individual metabolisms, COBRA community models have the potential to integrate various omics data and sequences, which could guide simulations to more biologically relevant results.

1.4.3 Model validation

Experimental validation of community simulation results play an important role to assess the relevance of the model and guide novel discoveries. While community models can give interesting novel hypothesis with their simulation, particular attention should be paid to the biological relevance of the predictions. It is therefore important to relate and validate at least part of the simulations with experimental values that come from existing knowledge or direct experiments. Existing knowledge can be used to validate the predictions of community models, e.g., by comparing simulated with measured metabolite concentrations [175]. This will also help to asses the relevance of the community models and how they should be interpreted. Since models of the human gut microbiota can be quite complicated and extensive, it is also difficult to find appropriate data or design experiments that can be used for validation. Part of the simulation results will be thus novel hypothesis that can only be validated with new experiments. This can guide the targeted design of experimental studies, which becomes important in the field of the human gut microbiota to reduce the complexity to simple findings.

1.5 Conclusions and future perspectives

COBRA community modeling approaches are promising tools to give novel insights and hypotheses of microbial consortia in the human gut. These models allow to integrate data and go beyond the descriptive nature of high throughput data analyses to unravel potential mechanisms. The different modelling paradigms we discussed have different assumptions that need to be assessed before starting an analysis. Based on this assessment, the model that is least complicated and most explanatory for the specific research question should be chosen. It could also be fruitful to combine different approaches to see their consistency or potential differences for a specific problem. It would be for example interesting to model single isolated species and communities separately to find beneficial or detrimental effects on the single species level. Further comparisons can be made between network based and modeling approaches: Can, for example, the negative and positive interactions in a co-existence correlation network be explained by the underlying metabolic interactions? Comparisons between algorithms could give further hints on potential experiments that can be performed for validation. This can drive further research on human microbiota, mechanistically predicting different treatments for various diseases.

1.6 Scope and aim of the thesis

The work in this thesis was aimed to investigate the metabolic potential of microbial communities in the human gut with constraint based modeling. This aim was divided into three main projects. First, genome scale metabolic models were automatically reconstructed and analyzed to assess the functional diversity of gut microbes, which demonstrated that each microbe has a particular metabolic repertoire. Second, a community modeling approach was developed to simulate the ecological interactions and metabolic exchange between these diverse microbes. This approach was then applied to small microbial community to model the effect of metabolite resources on the community structure. In the third and last project, the developed modeling approach was applied to human data by integrating metagenomic information with personalized community models, to simulate the differences between healthy controls and patients with inflammatory bowel disease. The constructed individualized com-

munity models were then used to predict potential treatments that improve the patients' profile in comparison to healthy controls. Below are short descriptions of each thesis chapter with the author contributions.

Chapter 2: Phenotypic differentiation of gastrointestinal microbes is reflected in their encoded metabolic repertoires

Chapter 2 describes the automatic reconstruction of 301 representative human gut microbes and the analysis of functional and phylogentic diversity of these gut microbes. The chapter is a full reprint of the paper published in BMC Microbiome on November 2015 [9].

Contributions

Eugen Bauer (EB), Ines Thiele (IT), and Paul Wilmes (PW) designed the study. EB and IT reconstructed the metabolic models and performed the analysis. Cedric C. Laczny performed the phylogenetic analysis. Stefania Magnusdottir collected phenotypic information and translated the reaction abbreviations. All authors edited and approved the final manuscript.

Chapter 3: BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities

Chapter 3 describes the development of a community modeling approach and its application to a simplified intestinal community of seven species. The chapter is a full reprint of the paper published in PLoS Computational Biology on May 2017 [10].

Contributions

EB, Johannes Zimmermann (JZ), IT, and Christoph Kaleta (CK) conceptualized the study. EB and JZ developed the methodology and software, performed the analysis, data curation, validation and visualization. EB, JZ, and Federico Baldini performed the analysis. IT and

CZ provided resources and project administration. All authors edited and approved the final manuscript.

Chapter 4: From metagenomic data to personalized in silico microbiotas: Predicting dietary supplements for Crohn's disease

Chapter 4 describes the construction of individualized community models based on healthy controls and patients with Crohn's disease to predict potential dietary treatments. The simulation results were used to first find differences between healthy controls and patients, which were partially validated with experimental knowledge. Then, the metabolic modeling approach was used to predict dietary treatments that could revert these differences. The chapter is currently a manuscript in preparation.

Contributions

EB and IT designed the study. EB performed the data integration and simulation of the personalized community models. EB and IT performed the analysis and writing of the manuscript.

Chapter 5: Concluding remarks

Chapter 5 is the personal thesis conclusion of the author with outlook to further studies.

Contributions

The text was written in full by EB.

Chapter 2

Phenotypic differentiation of gastrointestinal microbes is reflected in their encoded metabolic repertoires

Bauer, E., Laczny, C. C., Magnusdottir, S., Wilmes, P., and Thiele, I. (2015). Phenotypic differentiation of gastrointestinal microbes is reflected in their encoded metabolic repertoires. Microbiome, 3(1), 55. DOI:10.1186/s40168-015-0121-6

Abstract

The human gastrointestinal tract harbors a diverse microbial community, in which metabolic phenotypes play important roles for the human host. Recent developments in meta-omics attempt to unravel metabolic roles of microbes by linking genotypic and phenotypic characteristics. This connection, however, still remains poorly understood with respect to its evolutionary and ecological context. We generated automatically refined draft genome-scale metabolic models of 301 representative intestinal microbes in silico. We applied a combination of unsupervised machine-learning and systems biology techniques to study individual and global differences in genomic content and inferred metabolic capabilities. Based on the global metabolic differences, we found that energy metabolism and membrane synthesis play important roles in delineating different taxonomic groups. Furthermore, we found an exponential relationship between phylogeny and the reaction composition, meaning that closely

related microbes of the same genus can exhibit pronounced differences with respect to their metabolic capabilities while at the family level only marginal metabolic differences can be observed. This finding was further substantiated by the metabolic divergence within different genera. In particular, we could distinguish three sub-type clusters based on membrane and energy metabolism within the *Lactobacilli* as well as two clusters within the *Bifidobacteria* and *Bacteroides*. We demonstrate that phenotypic differentiation within closely related species could be explained by their metabolic repertoire rather than their phylogenetic relationships. These results have important implications in our understanding of the ecological and evolutionary complexity of the human gastrointestinal microbiome.

2.1 Introduction

Recent advances in sequencing technologies have greatly improved our knowledge about the metabolic complexity of the human microbiome and provide novel approaches to identify beneficial microbes [159]. In particular, sequencing the (ideally) entire genomic content (i.e., metagenomic sequencing) of the intestinal microbiota has allowed the establishment of a catalog of main groups of microorganisms present in the gastrointestinal tract and potential metabolic pathways [158] by avoiding culturing and isolation of individual microbial organisms. In this respect, endeavors of the human microbiome project [199] and the MetaHIT consortium [130] aim at establishing comprehensive data-sets of metagenomic content, metabolic functions, and taxonomic compositions within human individuals as well as the isolation and sequencing of numerous microbial taxa.

Despite these efforts, however, we are still lacking a comprehensive mechanistic understanding of the intestinal microbiota. One major hurdle in achieving this goal is the lack of organismal system boundaries, enabling us to associate the presence of metabolic pathways in the microbiome with a specific bacterium. Inferring metabolic roles by taxonomic classification alone is difficult because phylogenetically closely related organisms might be very different in their metabolism [135]. It may be therefore challenging to associate functional roles to entire taxonomic groups [6] to conjecture the biological relevance of intestinal bacteria. For instance, members of the same genus, or even of the same species, can be both probiotic and pathogenic [141], indicating a differential strain-specific adaptation. In this context, nutrient utilization can be a strong determinant for the adaptation to varying environments, since it can give a competitive advantage to other organisms that are metabolically less versatile. Thus, having additional metabolic functions can aid microbes in occupying further niches within the human gut. Accordingly, the functional consequences for the host change.

Current developments in systems biology allow the modeling of microbial metabolism to gain a mechanistic insight into the relationship between genotype and phenotype [167]. Genome-scale metabolic reconstructions form the basis of such modeling efforts. A reconstruction is assembled based on the genomic sequences as well as biochemical and phenotypic data of a target organism, and accounts for metabolic genes, enzymes, and their

associated reactions [145, 192]. Genome-scale metabolic reconstructions serve as a blueprint for condition-specific metabolic models [145, 192], which are obtained by the application of constraints, such as known nutrient uptake rates. The reconstruction process often includes a gap-filling procedure [103, 194], in which additional reactions are included to better model biologically relevant phenotypes, such as the formation of all known biomass precursors [160]. Metabolic models can be studied using a variety of mathematical methods [112]. One frequently used approach is flux balance analysis, which is applied to investigate a functional steady-state flux distribution of the modeled system, while maximizing (or minimizing) a particular cellular objective (e.g., production of biomass precursors) [145]. This modeling approach has been used to investigate nutrient requirements [188], gene essentialities [49], and metabolic interactions [80] for organisms of interest, thereby providing new insights into phenotypic and metabolic properties. The reconstruction process relies on the availability of detailed phenotypic data for the target organism [192], which is usually not available for many of the commonly found microbes in the human gut [159, 158]. To obtain representative metabolic reconstructions for these less well-studied organisms, automatic tools have been developed in recent years, such as the Model SEED platform [84], to provide a valuable starting point for metabolic modeling. In fact, draft reconstructions have been used to generate hypotheses about the target organisms with subsequent experimental validation, leading to the refinement of the metabolic reconstruction [160, 163, 144, 123].

In this study, we generated automatically refined draft genome-scale metabolic models of 301 representative intestinal microbes in silico based on whole genome sequences of the human microbiome project using an established approach [84]. We applied a combination of unsupervised machine-learning and computational modeling techniques to study individual and global differences of the metabolic models and the original genomes. Our key results include: i) divergent reactions involved in energy metabolism and membrane synthesis which are most relevant to discriminate different phylogenetic groups, ii) a linear relationship between differences in metabolic reaction potential and essential nutrients determined by flux balance analysis which indicates that the phenotype is directly correlated to the metabolic repertoire, iii) differences in metabolic reaction potential and phylogeny which exhibit an exponential relationship, suggesting an explanation as to why closely related microbes can be very different in their metabolic traits while at less-resolved phylogenetic distances only

marginal differences in metabolic diversity can be observed, iv) local differences in pathway presence which can be used to further distinguish representatives of *Lactobacillus*, *Bifidobacteria*, and *Bacteroides*. In summary, we demonstrate the importance of the metabolic repertoire of microbes to predict their phenotypic behavior in an ecological and evolutionary context.

2.2 Methods

2.2.1 Metabolic model selection, construction, and refinement

We selected a set of 301 microbes (Supplementary Table A.2) representing species present in the normal gut microbiota of healthy individuals, according to previous studies [159, 158]. We retrieved the genome sequences as well as additional information about the sequencing status, oxygen requirement, taxonomic placement, and phenotype from the integrated microbial genome database [126]. The completeness and possible genomic contamination by other microorganisms of the individual 301 genomes was assessed using a collection of 107 universal, single-copy genes [45]. The genomic sequences were uploaded for gene annotation to the RAST server [7] using default parameters. Draft metabolic reconstructions were then built with these genome annotations using the Model SEED pipeline [84]. To ensure, that the metabolic models are able to grow under anaerobic conditions, which are prevalent in their natural ecosystem, we modified, if necessary, one to five reactions to enable anaerobic growth. The reactions modified for each model are listed in Supplementary Table A.1. For descriptive purposes, reactions in the metabolic models were translated into our in-house metabolite and reaction database. The original SEED reaction nomenclature was maintained for the growth simulation. All refined draft metabolic models are publically available in their Matlab format at `http://thielelab.eu/in_silico_models`.

2.2.2 Growth simulation

To compute different growth conditions, the metabolic reconstructions were subjected to flux balance analyses [145] with the COBRA Matlab toolbox [167] using IBM ILOG cplex as the linear programming solver (IBM, Inc.). Briefly, genome-scale metabolic models were

represented as a stoichiometric matrix S, which encodes information about the mass balance of the complete set of enzymatic and transport reactions as well as a biomass reaction. The biomass reaction was retrieved from the metabolic reconstructions and represents the production of cellular building blocks (e.g., cofactors, amino acids, and lipids). Based on the stoichiometry, we could distinguish in our set of models 17 distinct biomass reactions, and based on the qualitative presence of compounds, we could distinguish 6 types of distinct biomass reactions (Supplementary Table A.2). Hence, the automatically included biomass reactions from the Model SEED pipeline are different and therefore reflect different precursor needs of the considered microbes. Given this reaction as an objective for the biological system, the metabolic fluxes of all reactions in steady-state maximizing growth can be determined by defining an optimization problem as follows:

$$\begin{gathered}\text{maximize } v_B \\ \text{subject to: } S \times v = 0 \\ v_{i,min} \leq v_i \leq v_{i,max}\,, \forall i \in n \text{ reactions}\end{gathered}$$

With v_b as the flux through the biomass objective function, v as the vector of all reaction fluxes, $v_{i,min}$ as the minimal flux capacity of reaction i and $v_{i,max}$ as the maximal flux capacity of reaction i. The solution (metabolic fluxes of all reactions) of this optimization problem can be obtained using linear programming. The flux through the biomass reactions can be interpreted as the growth rate of the microbe model. By setting the constraints $v_{i,min}$ and $v_{i,max}$ of exchange (transport) reactions, varying growth conditions can be simulated. Throughout this study, the maximal uptake was constrained to 10 mmol/gDW/h to estimate natural occurring conditions. The maximal achievable growth rate was calculated under these conditions by assuming that all exchange reactions are potentially active (equivalent to rich medium condition). Additionally, the absence of a particular metabolite in the medium was simulated by setting its minimal and maximal exchange reaction constraints ($v_{i,min}$ and $v_{i,max}$) to 0 mmol/gDW/h. By the iterative removal of each metabolite individually from the rich medium for each microbe model, different growth conditions were simulated. Essential nutrients were defined by growth rates smaller than 0.05 h^{-1} after removal from the medium. This cutoff was based on the estimated growth rate of microbes within the mammalian gut [64]; however, all calculated smaller growth rates were below 0.0001 h^{-1} and thus negligible.

2.2.3 Data mining of metabolic and genomic information

To assess the differences between the individual microbes, we used the reaction content and essential nutrients as well as COG functions and Pfam domains. The reaction content was based on the metabolic models obtained from Model SEED [146], whereas the COG functions and Pfam domains were obtained from the integrated microbial genomes database [126]. For each microbe, the presence and absence of reactions, essential nutrients, and functions were assessed in relation to the union of all metabolic reconstruction and genome annotations, respectively. The resulting binary vector b was then analyzed between species i and j with the Jaccard Index as:

$$\frac{b_i \bigcap b_j}{b_i \bigcup b_j}$$

to calculate the metabolic proximity according to [128]. Based on the obtained distance matrix of the reaction content, we used principle coordinate analysis [70] and t-SNE [202] for reducing the dimensionality from 301 to 2. The two-dimensional embeddings were visualized by scatter plots. Using principle coordinate analysis, we analyzed reaction differences between the metabolic models on a global scale by correlating each reaction to the principle coordinates and subsequently selecting the 200 reactions with the highest correlation (Supplementary Table A.3). The t-SNE-based visualization was used to identify local differences, with a detailed analysis of cluster structures within the genera *Lactobacillus*, *Bifidobacteria*, and *Bacteroides*. The reaction set differences between the determined sub-types of these genera were then used to identify type specific pathways.

2.2.4 Phylogenetic analysis

In addition to the determined metabolic difference, we used the phylogenetic relationships between the microbes as a measure of divergence. The phylogeny was computed with PhyloPhlAn, which uses a set of around 400 protein-coding genes for the phylogenetic placement [171]. In addition to the 301 bacterial genomes, the genomes of the archaea *Methanobrevibacter smithii* ATCC35061 and *Methanosphaera stadtmanae* DSM 3091 were used as an out-group to root the phylogenetic tree (Supplementary Figure A.2). The resulting phylogenetic tree was visualized using EvolView [219]. The phylogenetic difference between

the different bacteria was computed using the cophenetic distance based on the rooted tree [181].

2.2.5 Correlation between phylogeny, metabolic repertoire and essential nutrients

We determined the relationship between the metabolic repertoire of the models and the phylogenetic distance as well as its relation to the predicted essential nutrients by representing the phylogenetic distance as a function of the metabolic distance. We fitted different regression functions and found an exponential model defined by:

$$y = 10^{(\alpha+\beta x)}$$

to be the most suitable for explaining the relationship between metabolic distance x and phylogenetic distance y. For the relationship between essential nutrient difference z and metabolic difference x, we found a linear model defined by:

$$z = \alpha + \beta x$$

to be the best fit. We complemented the exponential model with the Spearman correlation and the linear model with the Pearson correlation as a measure of association between the variables. The goodness of fit measures for the different models and subsets of the data can be found in Table 2.2. The fitted parameters α and β for all plots in Figure 4 can be found in Supplementary Table A.7.

2.3 Results and discussion

2.3.1 Selected microbes as a model for the human gut microbiota

In order to answer ecological and evolutionary questions relevant for human health and disease, we selected 301 commonly found gut microbes based on their reported occurrence in the healthy gut microbiome [159, 158] and the availability of sequenced isolate genomes (Figure 2.1). We used the Model SEED platform [84] to generate automated draft genome-scale metabolic reconstructions for each microbe. To enable growth under anaerobic conditions,

which are predominant in the human gut [51], we added specific reactions, if necessary (Supplementary Table A.1). A comparison of our draft reconstructions with a set of published manually refined high-quality metabolic reconstructions taken from [82] revealed that most of the metabolic functionalities were captured in the refined draft reconstructions (Supplementary Figure A.1). Reactions absent in the refined draft reconstruction belonged mostly to the category of transport and exchange reactions, whose addition requires experimental and physiological data, as substrate specificity and transport mechanism is difficult to automatically annotate in microbial genomes [109].

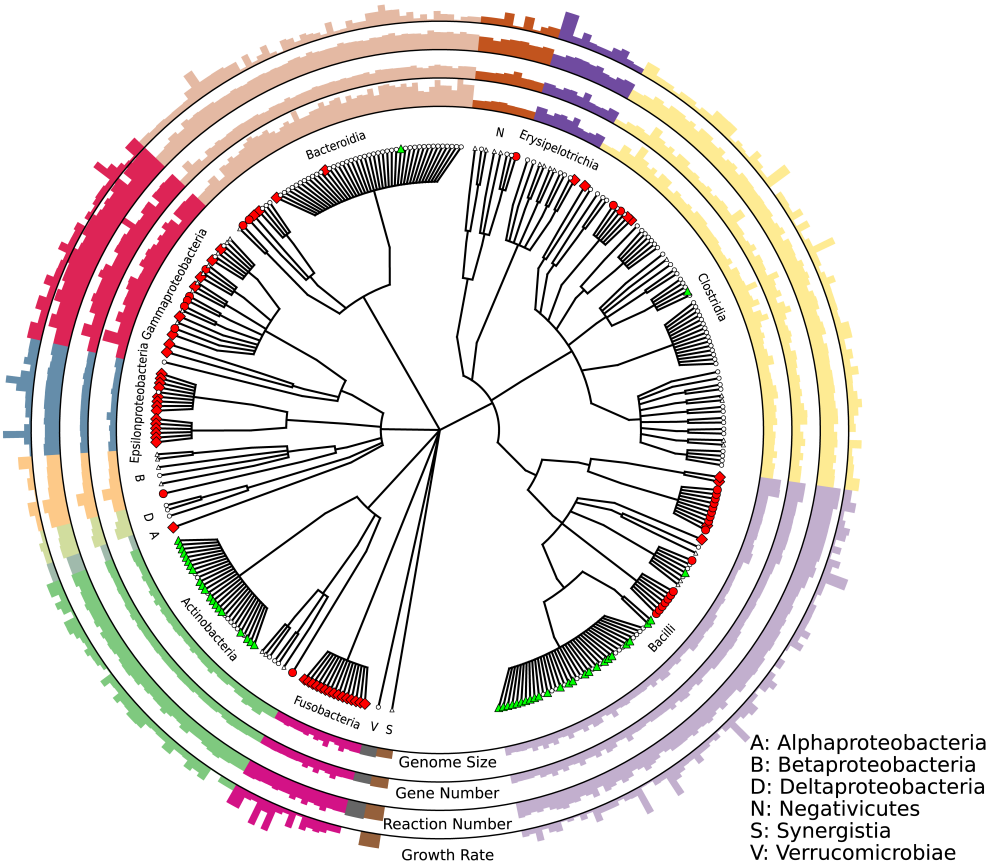

Figure 2.1: Phylogeny and individual statistics of the microbe selection. The cladogram shows the taxonomic relationships among the 301 microbes. In the four outer layers, the bars represent the relative individual genome size, number of genes, number of reactions, and in silico growth rate. The different colors represent the various bacterial classes. The leaf colors and shapes symbolize whether a microbe is a probiotic (green triangle), a pathogen (red diamond), an opportunistic pathogen (red circles), or a non-pathogenic bacterium (white triangles).

Our set of refined draft reconstructions captured a wide spectrum of different phyla

(Figure 2.1) with a taxonomic diversity similar to what is commonly observed in the human intestine [158]. The high diversity and proportion of microbes within the phyla Bacteroides, Proteobacteria, and Firmicutes (Figure 2.1) is concordant with observations in the human colon [182]. Moreover, by integrating information about pathogenic and beneficial traits of each microbe (Figure 2.1), we were able to associate these metabolic traits with the phenotype toward the host. As expected, a large proportion of probiotic *Lactobacillus* and *Bifidobacteria* could be found in the classes Bacilli and Actinobacteria, respectively (Figure 2.1 and Table 2.1). Additionally, known pathogenic organisms within the Proteobacteria, Fusobacteria, and Bacilli are also represented. Thus, our selection of bacteria provides an appropriate representation of microbial species, phenotypic traits, and metabolic processes present in the colon, the main site of microbial fermentation and interaction of microbes with the host [212].

Table 2.1: Genome and metabolic model statistics of the selected microbes.

Taxon	Number models	Average per taxonomic group				
		Genome (Mbp)	Genome complete[a]	Number genes	Number reactions	Gap-filled reactions[b]
All taxa	301	3.3	95%	702	875	3%
Class						
Bacilli	68	2.5	95%	656	860	4%
Clostridia	61	3.5	96%	735	839	1%
Bacteroidia	51	5.3	95%	764	908	1%
Actinobacteria	36	2.3	95%	514	727	10%
γ-Proteobacteria	25	4.6	95%	1127	1311	1%
Genus						
Lactobacillus	37	2.3	96%	597	798	4%
Bifidobacterium	29	2.2	95%	505	734	12%
Bacteroides	43	5.6	96%	774	909	1%

[a]Based on a selection of 107 essential genes [3].
[b]Based on the total number of reactions in the model. Gap-filling reactions were mostly added by the Model SEED platform.

Our analysis also included microbes with draft genomes (Supplementary Table A.2), requiring the assessment of the overall genome completeness and the potential impact on gene annotations and consequently on the generated metabolic reconstructions. The completeness and possible genomic contamination by other microorganisms of the individual 301 of the

individual genomes was assessed using a collection of 107 universal, single-copy genes [45, 157]. In our set of 301 genomes, the average estimated genome completeness was 95 % (Table 2.1). We further investigated the genome size and annotated genes among the 301 organisms (Table 2.1). Gammaproteobacteria had generally large genomes and a high number of annotated genes, while members of the order Bacteroidia had in general larger genomes but a lower number of annotated genes. This difference could be attributed to differences in annotation efficiencies, as Proteobacteria (in particular, gut specific *Escherichia* species) are very well-studied and thus have more homologous genes. Consequently, the number of reactions in the constructed metabolic models was higher and the number of reactions added via gap-filling lower. In contrast, we found a higher number of gap-filled reactions and a lower number of reactions in Actinobacteria (Table 2.1). This bias, which is well established for metagenomic analyses [25], is most likely the result of having less experimental data and validated gene annotation available for Actinobacteria. The presence of this apparent annotation bias underlines the limitation in current annotation techniques affecting particularly phylogenetically distant microbes [157, 25, 213] and highlights the need for more detailed experimental biochemical studies to elucidate gene functions in phyla distant to those containing model organisms [213].

2.3.2 Global reaction differences recapitulate conserved taxonomic patterns and phenotypes

To assess the differences within the metabolic reconstructions, we tested whether they could recapitulate the taxonomy of the studied microbes. We therefore computed a metabolic distance between the reconstructions based on the reaction presence [128] and subsequently used principle coordinate analysis (PCoA) [70]. This analysis revealed clusters, which correspond to known taxonomic groups (Figure 2.2). More specifically, with more than 30 % of explained variance, the first principle coordinate (Figure 2.2) was able to discriminate between Gram-negative and Gram-positive bacteria, which is in concordance to traditional measures of broad taxonomic groups, assigned based on the phylogeny of the 16S rRNA gene, the production of fatty acids, and corresponding membrane lipid composition [61]. In our PCoA (Figure 2.2), members of the class Negativicutes were closely associated with Gram-

negative bacteria rather than their phylogenetically close Gram-positive relatives, which is in accordance to their unusual membrane composition including two membrane layers [125].

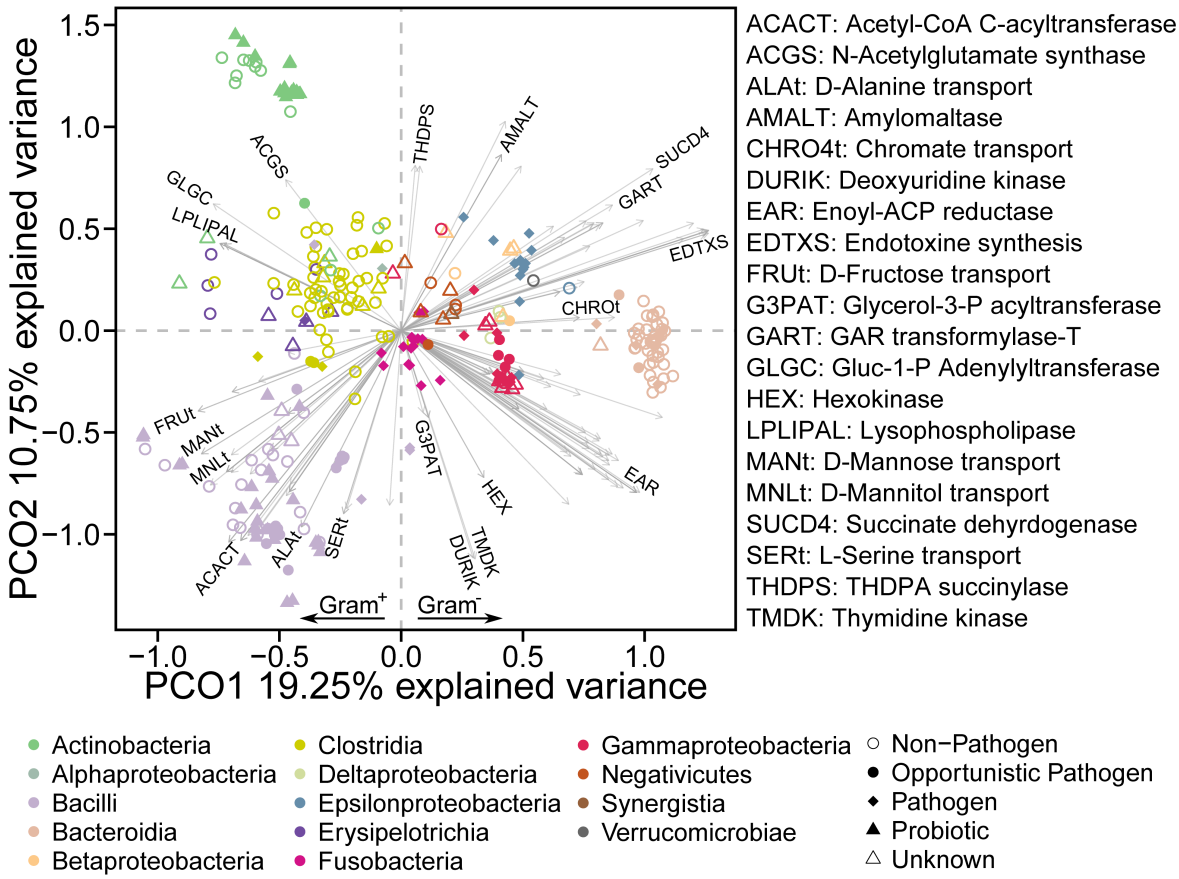

Figure 2.2: Global differences within metabolic models and their most divergent reactions. Biplot of the principle coordinate analysis based on the metabolic distance determined by the presence/absence of specific reactions in the metabolic models. Taxonomic groups are represented by different colors. The 200 reactions most associated with the point separation are indicated as arrows pointing from the coordinate origin to the contributing direction. The arrow shading represents reactions overlapping in their direction of contribution. The complete set of 2272 reactions sorted by their relevance can be found in Supplementary Table A.3.

The separation between Gammaproteobacteria and Actinobacteria highlights that our reconstructions captured taxa-specific metabolic features, despite the mentioned annotation bias. Furthermore, Clostridia species showed a high metabolic diversity and overlapped with clusters of other microbial taxa (Figure 2.2), which is consistent with the reported metabolic variety of these bacteria and their corresponding beneficial traits in the human gut [119]. Erysipelotrichia representatives are closely but nonetheless distinctly placed relative to the Clostridia in the 2D principle coordinate plot (Figure 2.2). Intriguingly, members of

Erysipelotrichia were formerly considered as Clostridia based on the phylogeny of marker genes [215] but then re-assigned to a novel class based on their phylogeny and membrane composition [207]. Similar to the Clostridia, Bacilli species were also widely spread in the 2D principle coordinate plot (Figure 2.2), reflecting their metabolic versatility [2]. In contrast, other taxa had more dense clusters, particularly Actinobacteria, reflecting more specialized roles of these bacteria, such as the conversion of polysaccharides [201].

Overall, we propose that metabolic reconstructions could be used, in addition to canonical approaches, to assist in the taxonomic definition of novel microbes and the re-assignment of already described microbes into better defined taxonomic groups. In particular, our approach has the advantage of considering functional characteristics, in contrast to methods solely relying on the presence and phylogeny of marker genes. As also pointed out by previous studies [220], functional repertoires can have a positive influence on the annotation quality of taxonomic groups. Ultimately, this could shed light onto the metabolic versatility of microbes in general or in specific habitats, such as the human gut.

2.3.3 Energy and membrane metabolism as markers for metabolic divergence

Following the broader characterization, we aimed to obtain a better understanding of the reactions driving the observed separation in the first two coordinates. The separation of taxonomic groups is due to reactions involved in membrane synthesis and central metabolism (Figure 2.2). In particular, different types of lysophosolipase reactions exhibit the highest explanatory power (Supplementary Table A.3). These reactions convert various phospholipid precursors (differing in their number of C-atoms) and have the same direction in the first principle coordinate, because all reactions can be carried out by single enzymes and are thus linearly dependent. Similarly, the amylomaltases catalyze multiple reactions differing in their substrates (Supplementary Table A.3). For the enoyl-ACP reductase, we found a variety of reactions with different directions toward the first principle coordinate (Figure 2.2). This variation in angle represents a potential variation in distinct yet convergent fatty acid synthesis processes involved in energy metabolism and known to be present in the human gut microbiome [57], thus contributing to the discrimination of the different

types of bacteria. This observation is consistent with the fact that fatty acid profiles have been used to characterize microbial communities before the advent of nucleic acid-based methods [76]. The synthesis of endotoxins was positively associated with the distribution of Gram-negative pathogenic species within the Proteobacteria and Fusobacteria, which is in accordance with previously reported correlations between various diseases and the abundance of Proteobacteria-producing endotoxins [196]. The transport and utilization of diverse carbohydrates involved in energy metabolism, such as mannitol, mannose, and fructose, were positively associated with the location of the Bacilli cluster in the 2D principle coordinate plot (Figure 2.2). This association highlights the variety of substrate consumption as represented by these reconstructions of microbial metabolism. In accordance with the literature, Bacilli are known to utilize a broad range of carbohydrates [94].

The differentiation of taxonomic groups based on reactions involved in energy and membrane metabolism may have important implications in understanding the evolution and heterogeneity of intestinal microbes. For instance, Gram-negative bacteria have been reported to change their membrane composition [77] in order to cope with environmental influences, such as antibiotics and human immune agents, many of which target bacterial membrane compounds [21]. Additionally, ecological changes within the microbial community [44] can provoke a differentiation in metabolic capabilities involved in energy metabolism leading to altered interactions of the community with the human host, supporting the observed high explanatory power of metabolic reactions toward cluster separation.

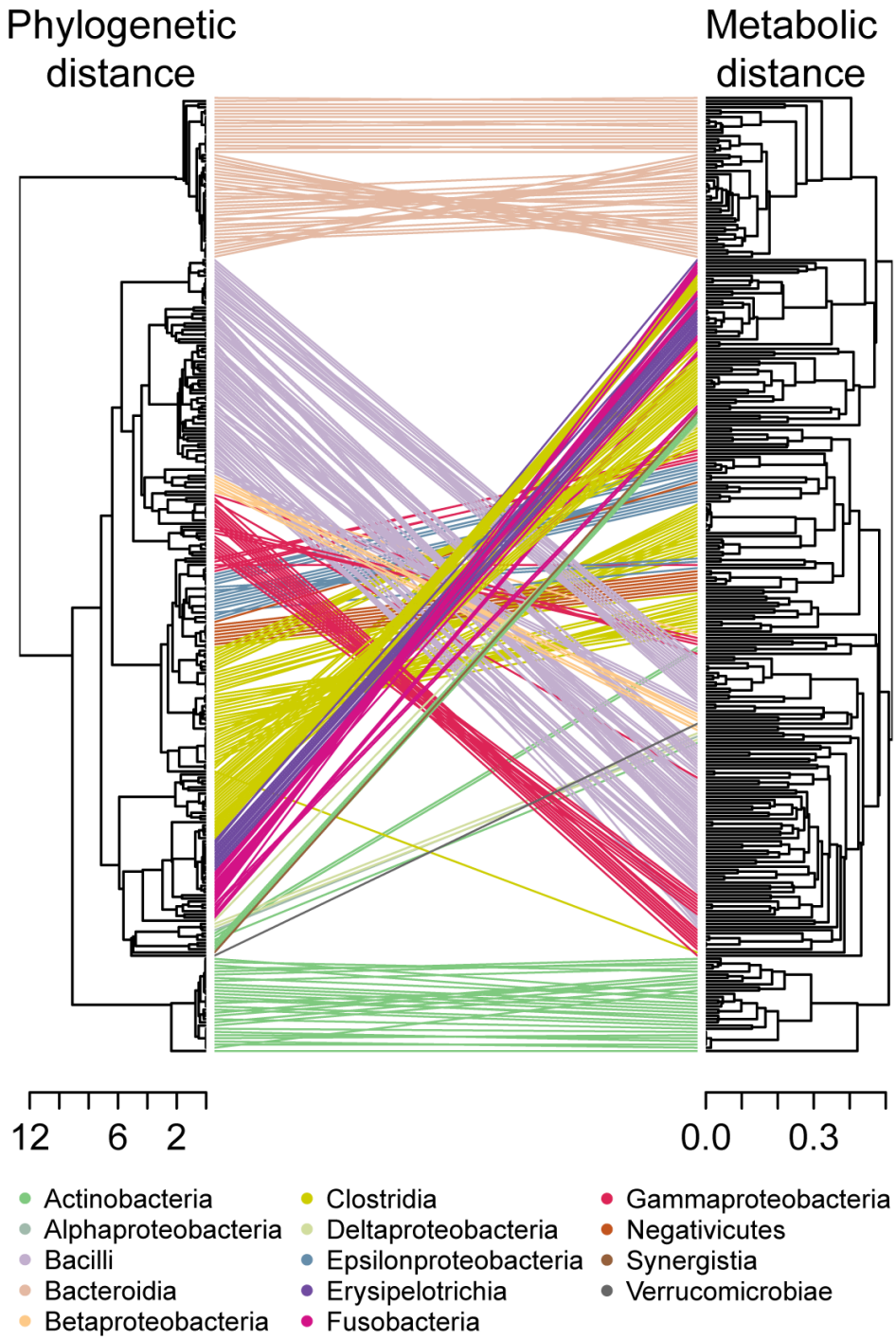

Figure 2.3: Tanglegram between the hierarchical clustering of the phylogenetic and metabolic distance. Tanglegram between the dendrograms of the reaction distance according to the presence of specific reactions and the phylogenetic distance according to the cophenetic distance of the maximum likelihood tree (rooted with two methanogenic archea) calculated from the sequence similarity of 400 selected essential genes. The dendrograms were calculated using hierarchical clustering with complete linkage. Lines connecting the same microbe are colored according to the taxonomic class.

2.3.4 The relationship between genotype, phenotype, and metabolic repertoire is non-linear

To further investigate the observed metabolic diversity (Figure 2.2) and its evolutionary basis, we computed the phylogenetic relationship between the 301 bacteria based on 400 protein-coding metabolic genes [171] using two methanogenic archaea as outgroups (Supplementary Figure A.2). On the basis of this rooted phylogenetic tree, we computed pairwise phylogenetic distances from the heights within the tree using the cophenetic distance [181]. While the clustering of this phylogenetic distance (Figure 2.3) recapitulated the original phylogeny (Supplementary Figure A.2), we additionally computed a genetic distance based on the 16S rRNA gene similarity of the microbes (Supplementary Figure A.3), to ensure that our observations were reproducible with other methods or markers. The pairwise distance based on the phylogenetic tree and the inferred presence of distinct reactions were overall congruent with each other (Figure 2.3). Interestingly, we identified an exponential relationship between phylogeny and metabolic repertoire (Figure 2.4), which is in accordance to a previous study based on genomic measures [216]. To exclude potential artifacts resulting from homology-based annotation methods (Model SEED) used for the generation of the metabolic reconstructions, we also determined the distance based on the presence of detected clusters of orthologous groups (COGs) [139] and Pfam protein domains [54]. These two measures also exhibited the same exponential relationship between metabolic repertoire and phylogeny (Figure 2.4). Importantly, this relationship indicates that closely related species can have an extremely divergent set of metabolic reactions, while at taxonomic ranks above the family level, only limited amounts of additional emergent features were observed. Since COG annotations and Pfam domains are prone to misclassification, we also included annotation measures with a higher quality, such as MetaCyc functionalities [26] as well as EC numbers (Supplementary Figure A.4) and observed a comparable exponential trend. Similar observations have been obtained in published experimental studies based on the phenotypic properties of different strains from the same genus or species [135, 141], underlining the biological relevance of our observations. In the context of a microbial community or biofilm, our observed relationship explains why closely related taxonomic groups (e.g., species of the same genus) are able to co-exist, while the overall consortium is limited in its metabolic potential [209]. In addition

to this result, we identified a linear relationship between the metabolic repertoire and the similarity of essential nutrients, which we calculated using flux balance analysis as a proxy for the metabolic phenotype (Figure 2.4b). These findings complement previous knowledge about the relationship between genotype and phenotype by Plata et al. [153]. Here, a similar exponential relation was observed between microbial phylogeny and varying growth conditions in selected genome-scale metabolic models, which were not directly associated with a specific habitat [153]. Additionally, this relationship has also been found with respect to the phenotypic similarity based on gene essentiality and synthetic lethal genes [153]. Taking into account that these latter measures have been based on flux balance analysis and are thus analogous to our results, we conclude that the observed patterns are generally applicable to bacteria. Furthermore, we argue that the metabolic network constituting of a set of reactions is appropriate to represent and explain a phenotype (Figure 2.4b). Assuming the metabolic repertoire as one of the major factors for the evolution of intestinal microbes, transfer of metabolic traits within different taxa may account for fast metabolic diversification of species and strains leading to niche partitioning. In fact, horizontal gene transfer has been shown to be enriched within organisms inhabiting the same environment, particularly, the human gut [176]. In addition to the results of Plata et al. [153], we propose the metabolic repertoire as one of the major factors influencing the phenotypic differentiation of human gut microbial communities. Still, the clear separation of taxonomic groups noted above (Figure 2.2) suggests that exchange of functionalities is limited to ensure a certain metabolic divergence within the whole microbiota to maintain functional diversity and limit competition between closely related organisms.

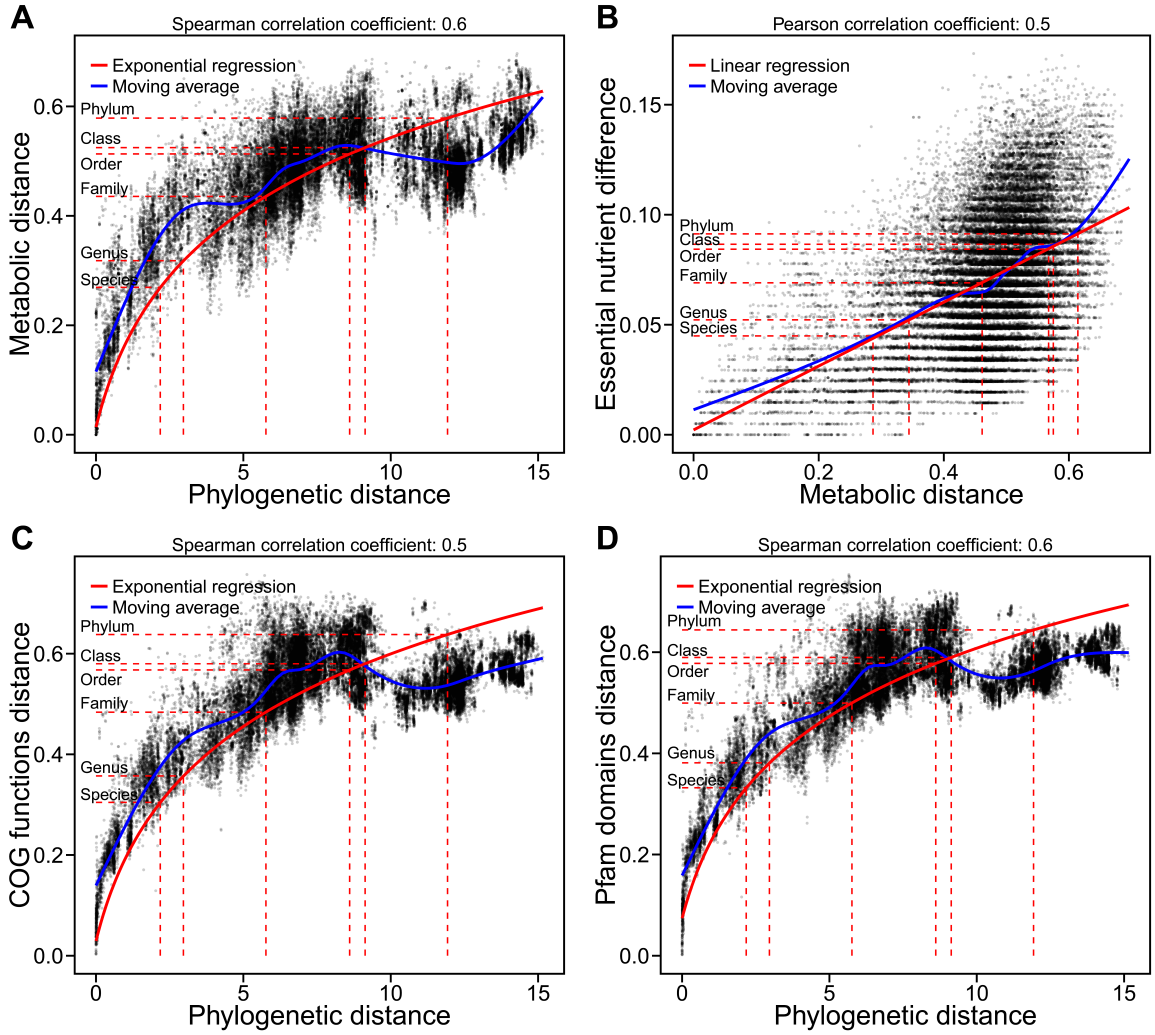

Figure 2.4: Relationship between reaction content, phylogeny, and phenotype. The metabolic distance was determined according to the presence of specific reactions in the model (a,b). COG (c) and Pfam (d) functional differences were assessed by comparing the presence/absence of COG functions and Pfam domains for all genomes, respectively. The phylogenetic distance is based on the cophenetic distance of the maximum likelihood tree (rooted with two methanogenic archea) calculated from the sequence similarity of 400 selected essential genes (a, c, d). The phenotype divergence was represented by the difference in essential nutrients, which were determined by removing the nutrient of interest from the in silico medium and subsequently checking for growth/no growth with flux balance analysis (b). The shading of the points (a-d) represents the density of all pairwise comparisons between the microbe models (n=45150). The blue line (a-d) represents the moving average over the data points. The goodness of fit for both regression models (a, b) can be found in Table2. The means of the phylogenetic (a, c, d) and metabolic distances (b) for each taxonomic category are indicated by dashed red lines.

2.3.5 The relationship between phylogeny, metabolic repertoire, and phenotype is taxon-dependent

To account for taxon-dependent differences between microorganisms (Table 2.1), we focused our analysis on model subsets of the five classes and the three genera with the highest number of representatives (Table 2.2). Additionally, this focus allows us to elucidate whether our results were dependent on our selection of microbes or could be expanded to other microbes not considered in this study. We found that the exponential relationship between phylogeny and metabolic repertoire as well as the linear relationship between nutrient essentiality and metabolic repertoire was apparent for most taxonomic groups (Table 2.2). However, we noticed differences within the taxa. In particular, there was a considerable exponential fit for all five major bacterial classes except for Clostridia, which could be explained by Clostridia's broad metabolic versatility and the corresponding difficulties in the taxonomic assignment within this class [33]. Our result is in accordance with the observed cluster variability of Clostridia when comparing the clustering of the metabolic and phylogenetic distance (principal coordinate analysis, Figure 2.3). When investigating individual genera, we detected a high correlation between essential nutrients and the metabolic repertoire of *Bifidobacteria*, whereas the correlation between their phylogeny and metabolic repertoire was less pronounced (Table 2.2, Figure 2.3). For members of the genus *Bacteroides*, the metabolic repertoire correlated strongly with their phylogeny (Figure 2.3), but only weakly with the essential nutrients (Table 2.2). Based on these results, we propose that the divergence within this genus can be explained by divergence in metabolic pathways relating to membrane synthesis (Figure 2.5) rather than energy metabolism and thus nutrient essentiality. For the *Lactobacillus* genus, we found a strong correlation between metabolic potential with both, phylogeny and essential nutrients. Within this genus, energy metabolism explained particular phenotypic divergences of species (Figure 2.5), which is consistent with the observed high correlation between reactions involved in nutrient uptake and the clustering of representatives of the Bacilli in the principal coordinate plot (Figure 2.2). Taken together, our results show a generality of the observed relationships between phylogeny, metabolic repertoire, and nutrient essentiality within and between taxonomic groups (Figure 2.4).

Table 2.2: Summary statistics of the relationship between reaction content, phylogeny, and essential nutrients.

	Reaction/phylogeny (exponential model[a])			Reaction/essential nutrients (linear model)		
Taxon considered	Spearman correlation	R^2	RMSE	Pearson correlation	R^2	RMSE
All taxa	0.59[b]	0.62[b]	0.11	0.48	0.23	0.03
Class						
Bacilli	0.68[b]	0.69[b]	0.08	0.58[b]	0.34	0.02
Clostridia	0.61[b]	0.39	0.12	0.67[b]	0.45	0.02
Bacteroidia	0.90[b]	0.80[b]	0.06	0.62[b]	0.38	0.02
Actinobacteria	0.80[b]	0.76[b]	0.16	0.86[b]	0.74	0.02
γ-Proteobacteria	0.78[b]	0.56[b]	0.11	0.72[b]	0.52[b]	0.01
Genus						
Lactobacillus	0.75[b]	0.70[b]	0.08	0.56[b]	0.31	0.03
Bifidobacterium	0.42	0.29	0.07	0.83[b]	0.69[b]	0.01
Bacteroides	0.81[b]	0.79[b]	0.05	0.33	0.11	0.02

[a]The exponential model was represented by a linear regression of semi-logarithmic transformed data.
[b]Values above 0.5.

2.3.6 Reaction differences reflect metabolic versatility among closely related microbes

To further investigate the metabolic divergence within closely related microbes of the same taxonomic group, we used t-distributed stochastic neighbor embedding (t-SNE) [202] for the two-dimensional visualization of the reaction similarities (Figure 2.5, see Supplementary Figure A.5 for point labels). t-SNE is a non-linear, non-parametric dimensionality reduction and has been used previously to reveal data-inherent cluster structures [4, 154, 105]. This method enabled us to identify fine-scale reaction differences, in addition to the principal coordinate analysis (Figure 2.2). Several distinct clusters were apparent and corresponded to the different bacterial classes (Figure 2.5). We further focused our analysis on the three most abundant genera in our model selection (Table 2.1). For *Lactobacillus*, we noted a widespread metabolic repertoire and thus a relatively large variability of members within this group (Figure 2.5). We identified three distinct subclusters (La1, La2, and La3) within this genus. While La2 showed major overlaps with the other two clusters, La1 and La3 were

distinct from each other. We investigated the differences in the reaction sets between the representatives of the different subclusters (Supplementary Table A.4). Based on the present reaction sets, La1 corresponds to obligate homofermentative La2 to facultative homofermentative and La3 to obligate heterofermentative pathways involved in the energy metabolism of lactic acid bacteria (Figure 2.5b). The pathway presence in the genomes explains why La2 overlaps with the other clusters, since the facultative homofermentative group (La2) shares reactions with the obligate homofermentative (La1) and heterofermentative group (La3) [94]. In agreement with the literature, these subclusters correspond to known divergent pathways involved in energy metabolism in Lactobacilli [2]. This distinction of biologically relevant phenotypic groups using predicted difference in metabolic reactions encouraged us to propose novel bacterial sub-types. Therefore, we confirmed for our choice of the number of subclusters by performing hierarchical clustering (Figure 2.3) to ensure that the subclusters were substantially different. For the *Bifidobacteria*, we propose two distinct subclusters (Bi1 and Bi2), which differed in the reactions involved in energy metabolism and membrane biosynthesis (Figure 2.5c). For the energy metabolism, numerous reactions involved in the uptake and utilization of diverse carbohydrates were observed for members of the subcluster Bi1 (Supplementary Table A.5), corresponding to known strain-specific differences within closely related *Bifidobacteria* [108]. Furthermore, we found reactions involved in the uptake and conversion of glucosamine to peptidoglycan, which could be associated with membrane composition in these two groups. To our knowledge, such pathway differentiation has not yet been proposed for *Bifidobacteria*. For the *Bacteroidia*, we could distinguish two subclusters (Ba1 and Ba2). The differences between these clusters can be attributed to the membrane biosynthesis (Figure 2.5d; Supplementary Table A.6). Members of Ba2 possess various pathway types leading to the production of varying phosphatidylglycerol compounds, whereas members of Ba1 can further process phosphatidylglycerol to myristic acid. This finding is of particular biological importance, when considering the virulence and signaling purposes of membrane lipids in *Bacteroides* species found in previous studies [5, 138], which links the phenotype to the synthesis of membrane compounds. Furthermore, since energy metabolism and substrate availability via the diet are major ecological driving forces within the human gut microbiota [57], the metabolic diversification of other closely related microbes, such as *Lactobacillus* spp. and *Bifidobacterium* spp., can be a necessary requirement to maintain

a stable coexistence with each other and the host. Considering that optimal conditions for metabolic cooperation are dependent on the similarity between the metabolic repertoires of several species [128], this pathway analysis approach could be used to estimate cooperative as well as competitive strategies. In particular, microorganisms tend to have a higher cooperativity if they are not too similar nor too different [28], indicating that members of the same taxon, but different subclusters (Figure 2.5b) might be able to co-exist, whereas functionally similar microbes may be more likely to compete with each other [209].

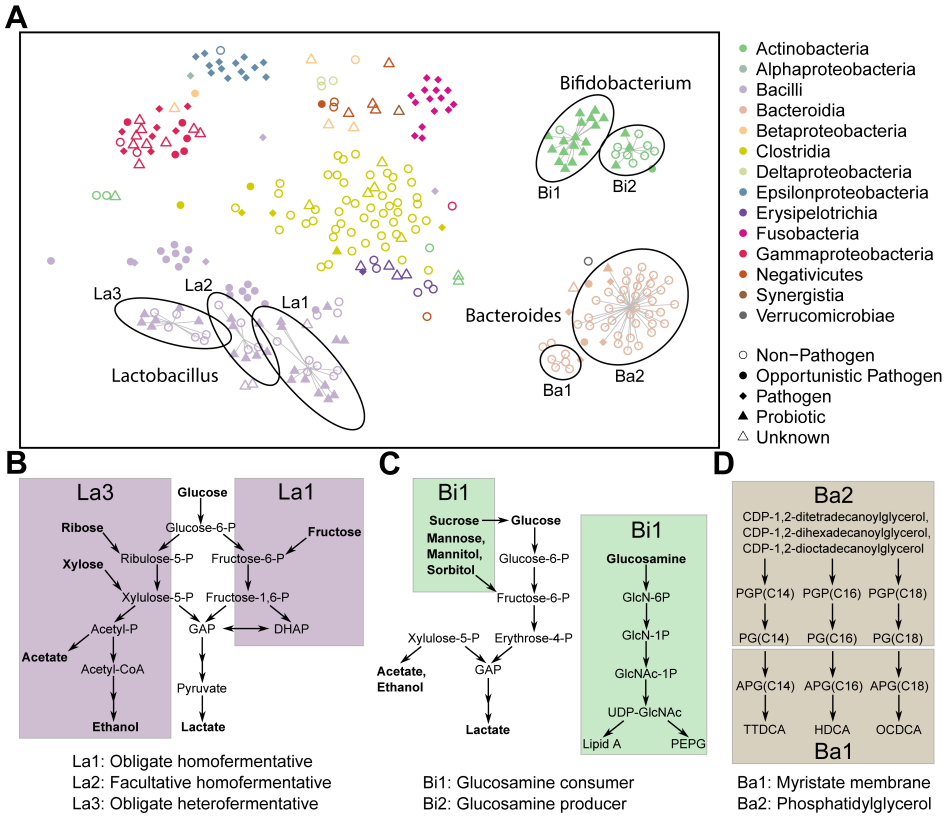

Figure 2.5: Local differences within metabolic models and their sub-type-specific pathways. The metabolic distance was determined according to the presence of specific reactions in the model (a). t-SNE was performed to obtain a low dimensional representation of the local differences within taxonomic groups which are represented by the different colors. Sub-types are defined based on hierarchical clustering of the reaction similarities. Members of one sub-type are connected with lines which originate from the cluster centroid. The ellipses represent confidence intervals of the clusters with a certainty of 95%. Distinguished pathways within sub-types include the genera *Lactobacillus* (b), *Bifidobacterium* (c), and *Bacteroides* (d). The pathways occurring in only one of the sub-types are framed by boxes carrying the corresponding cluster name. Reactions within pathways are represented by black arrows. GAP glycerol-3-phosphate, PEPG peptidoglycan, PGP phosphatidylglycerophosphate, PG phosphatidylglycerol, APG 1-Acyl-sn-glycero-3-phosphoglycerol, TTDCA tetradecanoate, HDCA heptadecanoate, OCDCA octadecanoate.

2.4 Conclusions

The requirement for a certain functional diversity to ensure a well-functioning cooperative intestinal microbiota is crucial to break down various complex dietary compounds and divide metabolic tasks among different community members [16]. Our results complement these ideas by investigating the metabolic divergence within a model microbiota, which can be primarily distinguished by reactions involved in energy and membrane metabolism. These capabilities play important roles in shaping the interface between host and symbionts, and thereby may lead to a deeper understanding in addition to metagenomic analyses in which all microbial functions are assessed [159]. Furthermore, the metabolic repertoire of microbes is proportional to their phenotypic properties, highlighting the importance of diversity in explaining the metabolic processes taking place within the human gut. In contrast to these properties, the metabolic repertoire exhibited an exponential relationship with phylogeny, underlining the challenges in inferring metabolic functions from phylogeny alone, in particular when using single gene-centric approaches such as via 16S rRNA gene amplicon sequencing. Moreover, this circumstance can be regarded as an important evolutionary and ecological feature of the microbiome; functional components constituting whole pathways can be very different within closely related species, whereas the metabolism in the overall metabolic repertoire is limited. In other words, by dividing the metabolic tasks between certain taxonomic groups, the microbiota can make efficient use out of a small set of functions thereby facilitating niche partitioning. This result has important implications when considering the overall species richness of the human gut microbiome in the context of different patients and diseases [106]. Further analyses could prove these concepts by modeling interactions within bacteria and the use of the here reconstructed and refined genome-scale metabolic models.

Acknowledgements

The authors are thankful to Mrs Almut Heinken for helping with the refinement of the draft metabolic models to account for anaerobic metabolisms and Dr. Dmitry Ravcheev for providing information about aerobic and anaerobic metabolisms. This study was supported by the ATTRACT programme grants (FNR/A12/01 and FNR/A09/03), the Aides à la Formation-

Recherche (FNR/6783162, FNR /4964712) grant from the Luxembourg National Research Fund (FNR), and a European Union Joint Programming in Neurodegenerative Diseases grant (INTER/JPND/12/01). None of the authors have any competing interests.

Chapter 3

BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities

Bauer, E., Zimmermann, J., Baldini, F., Thiele, I., and Kaleta, C. (2017). BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. PLOS Computational Biology, 13(5), e1005544. DOI:10.1371/journal.pcbi.1005544

Abstract

Recent advances focusing on the metabolic interactions within and between cellular populations have emphasized the importance of microbial communities for human health. Constraint-based modeling, with flux balance analysis in particular, has been established as a key approach for studying microbial metabolism, whereas individual-based modeling has been commonly used to study complex dynamics between interacting organisms. In this study, we combine both techniques into the R package BacArena (`https://cran.r-project.org/package=BacArena`) to generate novel biological insights into *Pseudomonas aeruginosa* biofilm formation as well as a seven species model community of the human gut. For our *P. aeruginosa* model, we found that cross-feeding of fermentation products cause a spatial differentiation of emerging metabolic phenotypes in the biofilm over time. In the human gut model community, we found that spatial gradients of mucus glycans

are important for niche formations which shape the overall community structure. Additionally, we could provide novel hypothesis concerning the metabolic interactions between the microbes. These results demonstrate the importance of spatial and temporal multi-scale modeling approaches such as BacArena.

3.1 Introduction

A major goal in microbial systems biology is to understand metabolic mechanisms underlying the emergence and organization of microbial communities [217]. Metabolic processes have been suggested to modulate and organize complex community structures by cross-feeding interactions (exchange of nutrients) [23, 186]. The human gut microbiota, for instance, consists of hundreds of species [47], whose compositions is strongly influenced by metabolic factors such as diet and microbial physiology [97]. Especially the metabolic interactions of multi-species communities within the gut have been found to support human well-being by the supplementation of nutrients via fermentation of otherwise indigestible dietary components [212]. One of the most important hallmarks in determining a healthy gut structure is the integrity of the mucus layer, which covers the epithelium, acts as a protective barrier against intruding pathogens, and enriches beneficial bacteria by providing nutritional compounds such as glycans [129]. Therefore, concentration gradients of substrates induce a spatial differentiation of the microbial community.

In biofilms, spatial concentration gradients of metabolites lead to a differential nutrient availability and therefore govern the distribution of species and phenotypes [23, 186]. Therefore, considering the processes that generate chemical gradients is essential when studying physiological heterogeneity in biofilms [186, 63, 149]. Individuals of the same or different species can support each other's growth by metabolic cross-feeding interactions [12]. Conversely, competition for nutrients can induce a division of metabolic tasks within the community which spatially differentiates the population in different sections, e.g. metabolically active and inactive microbe cells [116]. Such self-organizing processes have important implications in biomedical applications since single-species biofilms of pathogens are associated with a higher resistance against antibiotics [150] and thus obstructing potential treatments for diseases. In particular, most antibiotics are targeted at growing bacteria and not metabolically inactive dormant cell, which could re-initiate the biofilm after antibiotic treatment [116]. Furthermore, due to the physical structure of biofilms, antibiotics could poorly penetrate and often remain ineffective [185].

Constraint-based reconstruction and analysis (COBRA) is a key approach for the in silico study of microbial metabolism [112]. Metabolic reconstructions comprise the complete set

of biochemical reactions derived from a genome annotation in a stoichiometric accurate manner [192]. Through the application of specific constraints (e.g. nutrient availability) they can be converted into condition-specific models. With flux balance analysis (FBA), these models are used to optimize a given objective, such as the growth yield under a metabolic steady state [145]. To model metabolic interactions within microbial communities, different COBRA-based approaches have been developed [226]. First approaches modeled bacterial communities by combining the reconstructions of single microbes into a metabolic model, where metabolites can be exchanged and community growth is maximized using FBA [101, 81]. This concept has been recently expanded to allow integration of experimental data and modeling of distributed community growth [175]. Additional approaches have included temporal dynamics, in which microbial growth is simulated [221, 118, 224]. Recent advances incorporated spatial dynamics by enabling the distribution of microbes and metabolites, assuming homogeneous species populations [79]. Spatial environments were also used in an approach called MatNet [15] which combines FBA with individual-based modeling to simulate the metabolism of single species biofilms. Unlike population modeling, individual-based approaches allow to analyze populations as aggregations of autonomous individuals that interact by a set of rules. Accordingly, complex dynamics arise as emergent properties of locally interacting individuals [85, 74, 98].

In this study, we develop and apply BacArena, a community modeling tool which extends the integration of FBA and individual based modeling proposed by MatNet to model multi-species communities. Essentially, we model populations as aggregations of heterogeneous individuals that have their own metabolism and interact spatially as well as temporarily according to biologically relevant rules (e.g., movement, chemotaxis, and lysis). Furthermore, by modeling such metabolic heterogeneity, we can generate novel hypothesis concerning cross-feeding interactions within and between species. In particular, we applied BacArena to model *Pseudomonas aeruginosa* biofilm formation on the level of metabolic phenotypes. We could show how individuals are spatially arranged with different phenotypes according to nutrient availability. Furthermore, we identified phenotypes whose fermentation products contributed to growth of other biofilm members. In an application of a simplified human gut consortium consisting of seven species, we found that spatial gradients of mucus glycans are important to shape the community structure by forming a niche for glycan degrading

bacteria. Additionally, short chain fatty acids were exchanged between the community members and contributed to concentration levels which were similar to published experimental values. These results underline the increasing relevance of multi-scale modeling tools such as BacArena.

3.2 Methods

In principle, any genome-scale metabolic model in SBML or spreadsheet format can be imported and manipulated via *sybil* [62] and then directly integrated in BacArena. A hands-on tutorial for BacArena is available to illustrate specific use-cases and to get familiar with the code (Supplementary Note S1).

3.2.1 Concept and basic implementation of BacArena

We combine flux balance analysis (FBA) with individual based modeling. Each metabolic model belongs to an independent individual on a two-dimensional n × m grid environment (Figure 3.1) and acts according to biologically relevant rules (Table 3.1).

Table 3.1: List of rules implemented in BacArena and their corresponding references obtained from experimental studies.

Name	Description	Implementation	Ref
Metabolism	Computation of reactions speeds (fluxes)	Flux balance analysis (FBA)	[145]
Metabolism	Computing fluxes while minimizing enzyme usage	Parsimonious FBA	[111]
Kinetics	Defined metabolite uptake	Michaelis-Menten kinetics	[131]
Movement	Movement of individual cells	Random position change	[1]
Chemotaxis	Directed movement towards concentration gradient	Position change according to concentrations	[1]
Lysis	Cellular lysis after death	Secretion of biomass compounds	[132]
Growth	Biomass increase of each organism	Exponential and linear biomass increase	[136]
Death	Death of each organism	Organism death according to biomass threshold	[136]
Diffusion	Distribution of metabolites	Diffusion by partial differential equation	[184]

Consequently, FBA is a complex rule defined for an individual to compute the flux through all r biochemical reactions (flux vector $v \in \mathbb{R}^n$) by optimization of an objective function $c^T v$ (e.g., maximization of biomass yield). The corresponding linear programming problem can be written as follows:

$$\begin{aligned} \text{Maximize} \quad & c^T v \\ \text{Subject to} \quad & S \cdot v = 0 \\ & l \leq v \leq u \end{aligned}$$

where $S \in \mathbb{R}^{m \times n}$ denotes the stoichiometric matrix (m number of metabolites in an individual) and the vectors l and u represent the lower and upper bounds on n reactions respectively. The lower bounds of the external metabolite exchange are constrained according to the metabolite concentrations $[C_{i,j}] \in \mathbb{R}^m$ available at an individual's position (i, j) on the grid. All metabolites are initialized according to a initial concentration. Computed fluxes update the concentrations in every time step. Concentrations could be used as flux constraints because they represent the availability of the metabolites in the environment and therefore represent the uptake limit. Alternatively, if kinetic parameters are defined by the user, the lower bounds can be constrained according to Michaelis-Menten kinetics

$$l = \frac{v_{max} \cdot [C_{i,j}]}{K_M + [C_{i,j}]}$$

where v_{max} represents the maximal uptake rate and K_M the Michaelis-Menten constant, which can be obtained from public databases [170] or experimental data. The lower bound is constrained because exchange reactions are defined from the inside to the outside.

By default, FBA is used to calculate the metabolic fluxes given the metabolite concentrations of the local grid cells. Since most metabolic models are undetermined by having more reactions than metabolites, alternative optimal solutions (different flux distributions with the same objective value) occur during the simulations. To deal with this issue, we devised several alternatives to standard FBA calculations, which can be chosen by the user. For instance, parsimonious FBA can be used to minimize the total flux through all reactions of a metabolic model. In this case, the primary objective (e.g. biomass) is optimized first and afterwards a secondary objective (total flux) is minimized using the first optimal objective value as a

constraint. The second optimization acts as a proxy for minimal enzyme usage to simulate a more realistic behavior of cells in the exponential growth phase [111]. Additionally, the secondary objective can be chosen as a single reaction, which is picked randomly for each individual in each optimization, while enforcing the same biomass objective, pre-computed by FBA. The randomization of alternative optimal solutions can also be performed on the level of exchange reactions exclusively to get a better representation of secreted and consumed metabolites. The resulting flux distribution of the respective simulation strategy is then used to calculate and update the secretion or uptake for each individual in each simulation step. The linear programming problems can be solved using different solvers, such as GLPK [66], CLP [31], CPLEX [89], and Gurobi [78].

Based on the resulting FBA solution for each individual, exchange fluxes are used to update metabolite quantities $[C]$ in each grid cell. Moreover, the biomass B_t accumulated by an individual at time step t is updated according to an exponential growth model utilizing the optimal biomass yield $v_{biomass}$ computed by FBA with

$$B_{t+1} = B_t \cdot v_{biomass} + B_t$$

for each individual in each time step. The initial biomass (B_0) is selected according to the reported and experimentally determined median dry weight of one cell (Table 3.2). If multiple individuals are inserted in the environment, then a normally distributed random value is assigned to each individual, using the median and cell dry weight deviation (Table 3.2) as parameters for the normal distribution. When the total biomass of an individual reaches a duplication threshold, a daughter cell is spawned and placed at a free position in the Moore neighborhood (i.e. all surrounding grid positions in the direct neighborhood). The duplication threshold was chosen according to the experimentally determined maximum dry weight (Table 3.2), which represents the largest observed dry weight of one bacterial cell. To restrict growth to physiological feasible conditions, the accumulation of biomass is limited to 50% above the maximal cell weight. During optimization the upper bound of the objective function is set accordingly. If the biomass of an individual falls below a defined growth threshold, the corresponding individual dies (i.e. it is removed from the grid cell). If lysis is enabled, the biomass components can diffuse to neighboring grind cells.

The growth threshold was chosen according to the experimentally determined minimum dry weight (Table 3.2), which represents the smallest observed dry weight of one bacterial cell.

Movement is implemented as a random walk of individuals using unoccupied grid positions in the Moore neighborhood. Different movement velocities can be imposed by setting the number of grid positions to which an individual can move. Individuals can also perform chemotaxis by moving towards a concentration gradient of a particular metabolite of interest. Diffusion of the metabolite concentration $[C]$ in the two-dimensional x, y environment is implemented using Fick's second law of diffusion which in two dimensions reads

$$\frac{\partial [C]}{\partial t} = D \cdot \left(\frac{\partial^2 [C]}{\partial x^2} + \frac{\partial^2 [C]}{\partial y^2} \right)$$

where $D \in \mathbb{R}^s$ is a vector of diffusion constants. Zero-gradient boundary conditions are set to ensure mass conservation. The diffusion model is defined using the R package *ReacTran* [179] and is solved by the integrator *lsodes* (R package *deSolve* [180]). Additional diffusion functionalities, such as advection or different boundary conditions, are available and additional ones can be implemented with *ReacTran*.

To analyze population heterogeneity in terms of the metabolic turn-over, we defined metabolic phenotypes p by

$$p = \begin{cases} 1, & \text{if } v_{ex} > \quad \theta. \\ -1, & \text{if } v_{ex} < -\theta. \\ 0, & \text{otherwise.} \end{cases}$$

according to an adjustable threshold θ (default value is $\theta = 10^{-6}$) and considering the exchange reaction flux v_{ex} of each individual. The metabolic phenotypes represent the metabolic signature of all secreted and consumed metabolites for each individual. The metabolic phenotypes track the metabolism of each individual during each simulation step and thus indicate how each microbial cell changes the environment and interacts with other species. In addition, BacArena provides a range of different data analysis techniques within the R environment to investigate the emergence of complex phenotypes on the population level (see reference manual in Text S2).

Table 3.2: Default parameters of BacArena with references and the name of the variable set for the respective function.

Description	Function	Value	Unit	Bionumber [134]	Ref
Cell space occupation	Organism	4.42	μm^2	105026	[156]
Maximal dry weight	Organism	1.172	pg	106615	[117]
Minimal dry weight	Organism	0.083	pg	106615	[117]
Biomass decrease	Organism	0.210	pg	-	[117]
Median cell dry weight	Organism	0.489	pg	-	[117]
Dry weight deviation	Organism	0.132	pg	-	[117]
Oxygen diffusion (aqueous)	Substance	$20 \cdot 10^{-6}$	$cm^2\ s^{-1}$	104440	[184]
Glucose diffusion (aqueous)	Substance	$6.7 \cdot 10^{-6}$	$cm^2\ s^{-1}$	104089	[184]
Oxygen diffusion (biofilm)	Substance	$12 \cdot 10^{-6}$	$cm^2\ s^{-1}$	-	[184]
Glucose diffusion (biofilm)	Substance	$1.675 \cdot 10^{-6}$	$cm^2\ s^{-1}$	-	[184]
Glucose uptake Km	setKinetics	0.01	mM	-	[69]
Glucose uptake Vmax	setKinetics	7.56	$mmol g^{-1} h^{-1}$	-	[69]

3.2.2 Parameters, units, and integration of experimental data

BacArena permits fine tuning of simulations through adjustment of parameters which are incorporated in the different classes (Supplementary Figure B.1). The default parameters of BacArena are taken from various experimental data sets (Table 3.2). Based on the user defined length of the environment dimensions (in cm) and the number of grid cells, these parameters are automatically adjusted to represent physically meaningful results. Given the corresponding size of a grid cell (cm^2) and the occupied space of the organism of interest (μm^2), the maximal number of individuals per grid cell is computed and the maximum possible biomass per grid cell is calculated accordingly. Metabolite concentrations are integrated by converting molar concentrations (in mM) into metabolite amounts per grid cell based on the above defined geometry. Fluxes are calculated in $fmol \cdot (pg_{dryweight}\, h)^{-1}$.

3.2.3 Syntrophic two-species community model

Manual curated genome-scale metabolic models were retrieved for the hydrogen producing bacterium *Clostridium beijerinckii* [107] and the methanogenic archaeon *Methanosarcina barkeri* [67]. The *M. barkeri* model was modified to ensure methane production with hydrogen and carbon dioxide by blocking the uptake of acetate and only allowing unidirectional uptake of hydrogen, hydrogen sulfide, and sulfur trioxide. The *C. beijerinckii* model was modified to block the secretion of acetate in order to ensure hydrogen production. To model metabolic exchanges between the microbes and compare the results of BacArena, we performed the simulations with our method and COMETS [79]. For both methods, simulations were carried out on a 100 times 100 grid environment for 24 hours. In both setups, a minimal medium was added to the environment with 1 mmol of glucose per grid position, carbon dioxide, and several co-factors (4 aminobenzoate, cobalt, nicotinic acid, water, protons, ammonium, nickel, phosphate, sulfur trioxide, cysteine, and sulfate). To ensure the growth of *M. barkeri* before *C. beijerinckii* produces a sufficient concentration of hydrogen, an initial amount of 10^{-10} mmol hydrogen was added to each grid position. The diffusion of metabolites was calibrated to the standard diffusion of glucose (Table 3.2). For COMETS the diffusion was executed one time per iteration to create a similar setting as in BacArena.

3.2.4 *Pseudomonas aeruginosa* single-species biofilm model

Biofilm formation of *Pseudomonas aeruginosa* was simulated using the genome-scale reconstruction iMO1056 [142] retrieved from [15]. The reconstruction was modified to enable lactate fermentation (see Supplementary Note S1, Supplementary File S1). All growth parameters were set to default values (Table 3.2). The environment was initiated to represent one individual per grid cell and 100×100 grid cells, and therefore defining the spacial extent by $0.025mm \times 0.025mm$. Simulations were repeated ten times. For the starting condition, 900 individuals (9% inoculation) were placed into the center of the environment. Minimal medium, as described in [15], was used for each grid cell (Supplementary Table B.1). $50\mu M$ of glucose were added and all other metabolites of the minimal medium were initialized with a concentration of $100\mu M$. Glucose uptake of each individual (i.e. *P. aeruginosa* metabolic model) was constrained according to Michaelis-Menten kinetics based on published values

(Table 3.2). All remaining exchange reactions were unconstrained. Metabolites were allowed to diffuse freely with particular diffusion rates for gaseous and organic compounds in biofilms (Table 3.2). The simulation was performed for 48 time steps totaling to a simulation time of 2 days. The code and the model to reproduce the results of the simulations is provided in Supplementary File S1, Supplementary File S2. To model the influence of nitrate as additional electron acceptor, we used the results of the first 20 hours to resume the simulation after adding $0.01mM$ of nitrate. All simulations were performed using pFBA to generate the flux distributions of each individual.

3.2.5 Integrated multi-species model of the human gut

A model for the human gut was assembled using seven recently reconstructed genome-scale metabolic models of human gut bacteria [183]. In this study, the models were manually curated and checked using published experimental data. The bacterial species were selected according to their relevance and abundance within the human gut microbiota to represent a simplified human intestinal microbiota (SIHUMI) [11]. The following microbial reconstructions were used *Anaerostipes caccae* DSM 14662, *Bacteroides thetaiotaomicron* VPI-5482, *Blautia producta* DSM 2950, *Escherichia coli* str. K-12 substr. MG1655, *Clostridium ramosum* VPI 0427, DSM 1402, *Lactobacillus plantarum* subsp. plantarum ATCC 14917, *Bifidobacterium longum* NCC2705, and *Akkermansia muciniphila* ATCC BAA-835. The models used for the simulations are available on vmh.uni.lu as well as Supplementary File S2.

Growth parameters and movement were set to the default values (Table 3.2) and the environment was initialized with a 100×100 grid corresponding to a side length of $0.025mm$. Simulations were repeated five times, each time simulating $16h$ with time steps of $1h$. In a first condition, the intestinal lumen was initialized with 200 individuals of each species in an environment which was devoid of mucin glycans. All remaining metabolites were set to a concentration of $0.1\mu M$ except essential nutrients They were set to $1\mu M$ to ensure that all bacteria were able to grow. The essential metabolites were determined using flux variability analysis [122] on all unbounded exchange reactions for each metabolic model, while enforcing a minimal biomass rate of $0.01h^{-1}$. The exact diet definition with the predicted essential

metabolites can be found in Supplementary Table B.2. To investigate the importance of spatial concentration gradients, we devised a second condition, in which mucus glycans ($1\mu M$) were added as a linear gradient with decreasing concentrations from the bottom to the middle of the environment. Metabolites were allowed to diffuse according to diffusion rates for gaseous and organic compounds in aqueous solutions [184]. Mucin glycans were not allowed to diffuse, since they are known to be tightly associated with the epithelium in form of a mucous layer [177]. The code and the models to reproduce the results of the simulations are provided in Supplementary File S3, Supplementary File S4. All simulations were performed using pFBA to generate the flux distributions of each individual.

3.3 Results and discussion

With BacArena we provide a modular and extendable R package for modeling and analyzing microbial communities (Supplementary Note S1 and S2). In BacArena each organism is represented individually on a two-dimensional grid to model a spatial environment (Figure 3.1). Temporal dynamics are modeled by including time steps in which the state of each individual and the environment is updated. In each time step metabolites diffuse in the environment and can be exchanged between the individuals. Individuals can move to and duplicate within the neighboring grid positions. The metabolism of each individual is modeled by flux balance analysis on the underlying genome-scale metabolic model of the particular species. Using the biomass as an objective for the FBA and the metabolite concentrations in the corresponding grid position as constraints, the growth and metabolic turn over is determined. Accordingly, the duplication rate is obtained from the growth rate and the metabolite concentration is updated according to the secreted and consumed metabolites. Since a FBA is computed for each individual, every microbial cell can be heterogeneous in its metabolism and has therefore its own metabolic profile. These profiles are recorded as metabolic phenotypes in BacArena and can be used to infer cross-feeding interactions.

3.3.1 Comparison to other methods

Established methods in community modeling can be roughly divided into two groups: Equation based, continuous methods modeling populations (e.g. COMETS, dOptCom) and rule-

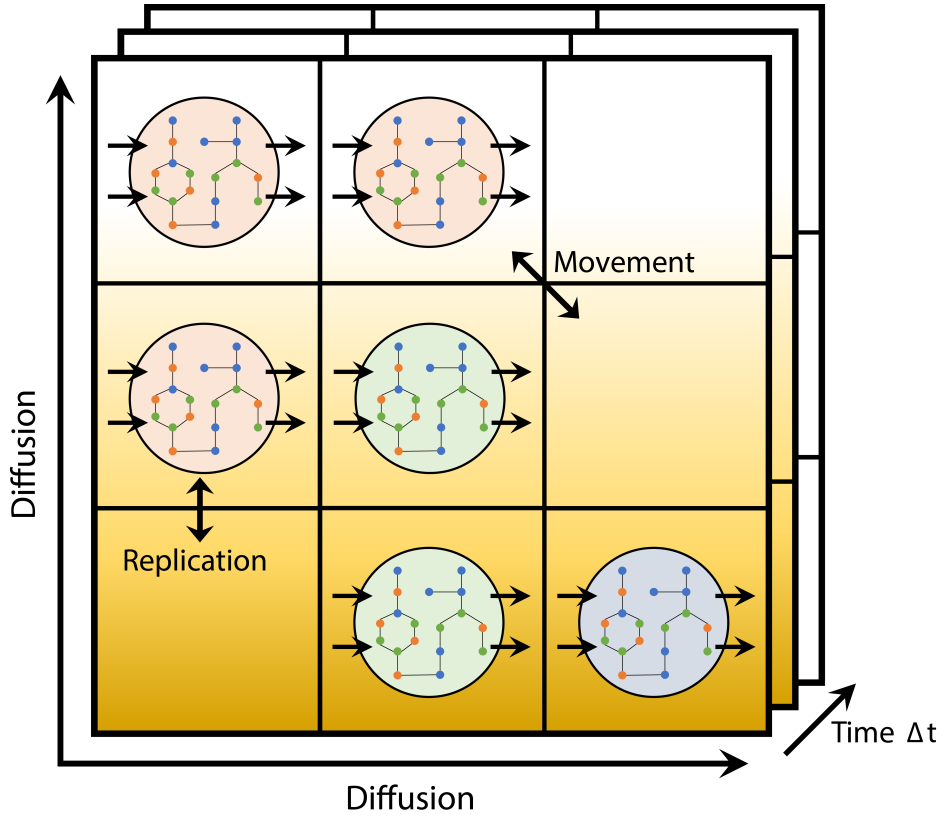

Figure 3.1: Schematic overview of BacArena. Microbial species are shown in different colors. Fluxes of exchange reactions are indicated as uni-directional arrows, movement and replication as bi-directional arrows.

based methods focusing on individuals (MatNet, BacArena) (Table 3.3).

BacArena extends the individual-based modeling approach of MatNet [15] to include more features (Table 3.3) and simulation of up to hundreds species (Figure 3.2). The runtime of BacArena simulations is linearly dependent on the number of individuals (Figure 3.2A) and increases till an addition of about 50 species (Figure 3.2B). Afterwards the runtime remains approximately stable because the diffusion of metabolites is computationally expensive and if including more than 50 species only few new metabolites need to be added. BacArena was developed to run efficiently even with large data sets due to R's capacity to integrate C++ code into time-consuming routines [48]. Additionally, computations can be executed in parallel to accelerate runtime.

To illustrate the difference between continuous and rule-based population modeling approaches, we compared BacArena and COMETS [79] in the context of a two-species syntrophic community of the methanogenic archeum *Methanosarcina barkeri* and the hydrogen

Table 3.3: Comparison of BacArena with other community modeling approaches involving metabolic models.

Method	Approach	Time	Kinetics	Space	Phen.	Parallel	GUI	Species
BacArena	FBA/ABM	✓	✓	✓	✓	✓		> 2
MatNet[15]	FBA/ABM	✓		✓			✓	1
COMETS[79]	dFBA	✓	✓	✓		✓	✓	> 2
dOptCom[224]	Multi-obj.	✓	✓					> 2
MCM[118]	dFBA	✓	✓				✓	> 2
DyMMM[221]	dFBA	✓	✓					2

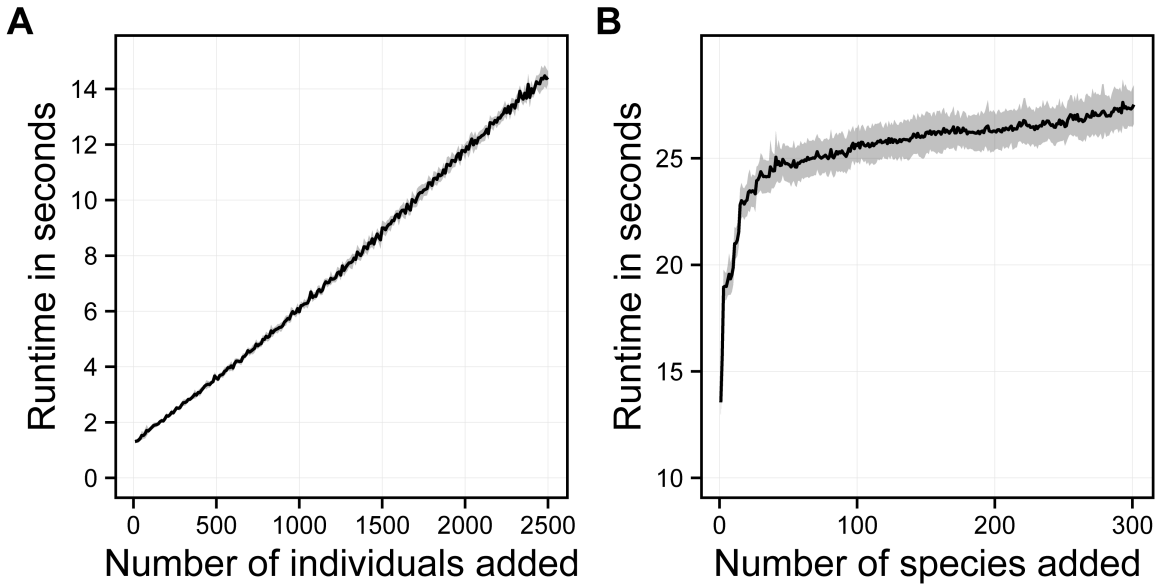

Figure 3.2: Runtime of BacArena in relation to the number of added individuals and species. A Runtime based on an example draft metabolic model (Clostridium sp. SY8519 model taken from [9]) with an increasing number of individuals added to an environment with a dimension of 50 times 50 grid cells. B Runtime based on an increasing number of species (301 draft metabolic models taken from [9]) added to an environment with a dimension of 50 times 50 grid cells and one simulation step. All simulations were run on a windows machine with 32GB of RAM and a 3.5GHz processor with four physical cores.

producing bacterium *Clostridium beijerinckii* (Figure 3.3). The hydrogen produced by *C. beijerinckii* is taken up as an electron donor by *M. barkeri* to reduce carbon dioxide to methane, which is secreted into the environment. This is in concordance with experimental knowledge, showing the metabolic exchange between hydrogen producing bacteria and methanogenic archaea [137]. Notably, COMETS and BacArena produce similar results in terms of these predicted cross-feeding interactions and are therefore consistent. Based on

the quantitative biomass production, both methods predict a smaller growth of *M.barkeri* compared to *C.beijerinckii*, however, the biomass production is higher in COMETS compared to BacArena. For the exponential phase of each simulation COMETS predicted a doubling time of 0.5h, BacArena predicted 1.1h, and the experimentally measured value is 4.3h [133]. The reason for this difference can be attributed to the underlying growth model of both methods. COMETS models colony growth as a 2D diffusion while BacArena models individual cell behavior and replication which causes the population to grow slower in the initial phase to reach a certain number of individuals. In BacArena populations consist of heterogeneous individuals (bottom-up) which have their own characteristics, e.g. movement and metabolic phenotypes. COMETS, on the other hand, is a top-down approach describing colonies on the population level (Figure 3.3). Both approaches differ concerning the representation of the spatial scale. In BacArena one individual is represented per grid position, whereas COMETS represents a population of multiple cells per position. Both, BacArena and COMETS, can predict heterogeneous growth rates according to spatial concentration gradients. By focusing on individuals, BacArena can be used to model additional heterogeneity of cells by accounting for their history and by integration of further rules such as cellular lysis. The explicit consideration of heterogeneous individuals has been regarded as especially helpful for addressing the complexity of biological systems, because local species interactions can represent biological systems more realistically [98, 68, 174]. In particular, the heterogenic movement in BacArena can be relevant when modeling an aqueous or viscose environment, such as the human gut, in which the movement is accelerated. Furthermore, by combining individual-based modeling with FBA, BacArena can model the metabolic state of each individual cell to investigate metabolic heterogeneity within a population of cells. This metabolic heterogeneity is captured by our definition of metabolic phenotypes, whose applicability and biological relevance we show in the next section on the basis of a biofilm model of *Pseudomonas aeruginosa*.

3.3.2 P. aeruginosa single-species biofilm model

To demonstrate the applicability of BacArena to biofilm formation, we constructed a single species biofilm model of *P. aeruginosa*. We used a glucose-minimal medium with oxygen

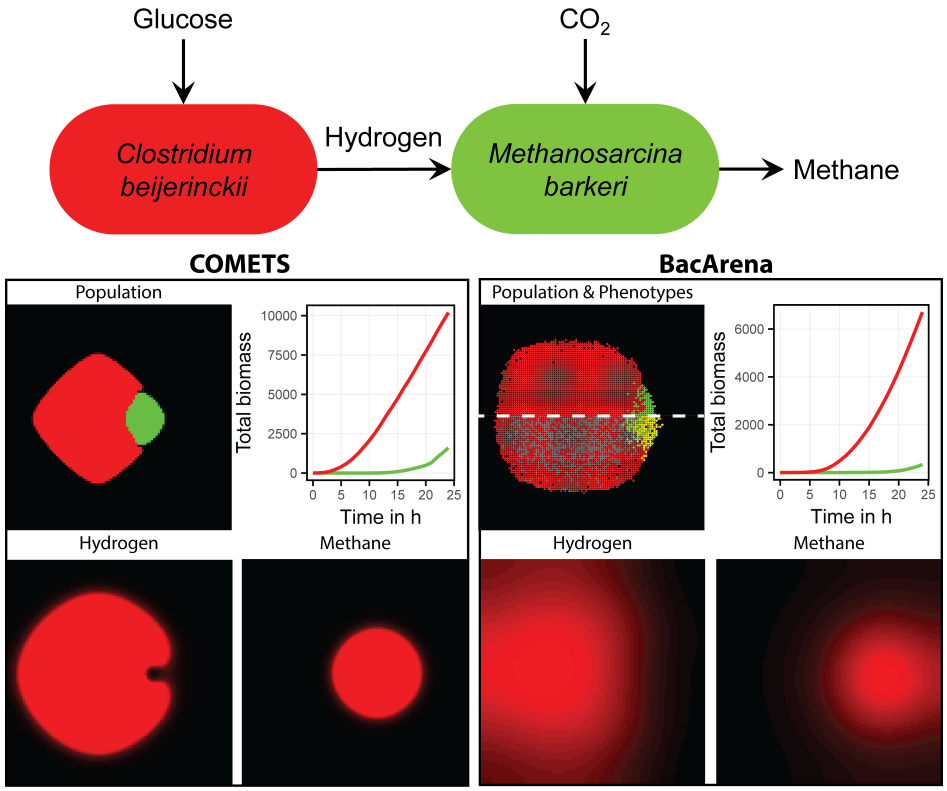

Figure 3.3: Comparison between COMETS and BacArena based on a simple two-species syntrophic community. The community is based on the published metabolic model for the hydrogen producing *Clostridium beijerinckii* [107] and the methanogen *Methanosarcina barkeri* [67]. Both simulations were carried out on a 100 times 100 grid environment. As initial concentrations, 1 mmol of glucose, carbon dioxide, and several co-factors were added per grid position. Grey cells in the phenotype plot of BacArena (lower half of the population plot) represent metabolically inactive cells.

as electron acceptor to investigate the metabolic behavior of individual cells of the biofilm community. We found spatial and temporal differences within the community which could be attributed to distinct emergent metabolic phenotypes. The observed phenotypes (P1-P9) were classified according to the usage or production of glucose, oxygen, acetate, succinate, and CO2 (Figure 3.4C). The phenotypes occurred in all replicate simulations (n = 10) with similar temporal dynamics (S2 Fig). Additionally, the phenotype appearance was stable with respect to variations in initial glucose and oxygen levels (S3 Text). Finally, we validated the growth model with experimental data [104] and correctly predicted a higher population size under rich conditions compared to a minimal medium (Figure 3.4E).

In the beginning of the simulation, we observed only a glucose oxidation phenotype (P3) that constituted the whole population (Figure 3.4A). After two hours the individuals

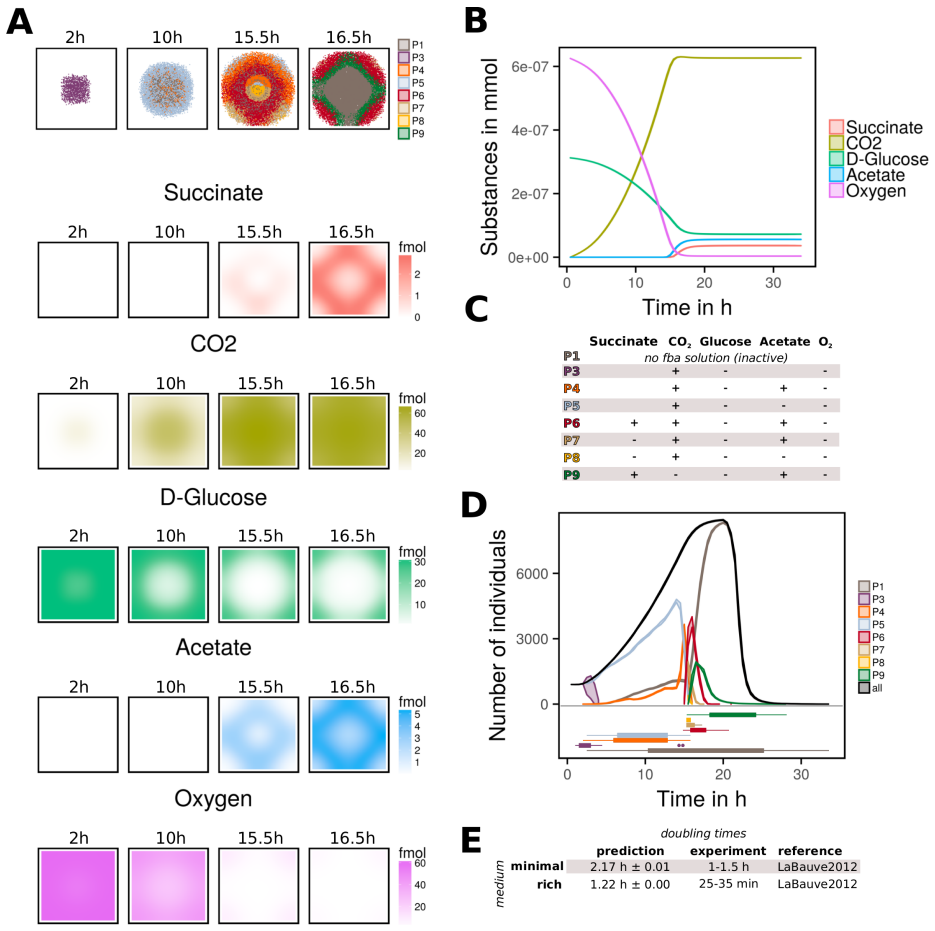

Figure 3.4: Single species biofilm model of *P. aeruginosa*. A Spatial distribution of individuals and key metabolites at different time points (2 h, 10 h, 15.5 h, 16.5 h). Different metabolic phenotypes are colored and represent community members with distinct production and consumption of metabolites. A metabolic inactive core was formed after 16 h and several fermentative phenotypes occurred in the outer layer of the biofilm (see subfigure C for description of phenotypes). Glucose and oxygen were consumed and CO2, acetate, succinate were produced. B Time curve of key Metabolites. Metabolites are given in mmol. Oxygen was consumed in total and some glucose remained in the end. Acetate and succinate levels increased after 15h. C Characterization of eight metabolic phenotypes (P1,P3-P9). Only phenotypes which occurred consistently in all replicates were considered. Therefore, P2 (growth with CO2 and acetate) and P10 (acetate and succinate production, glucose and oxygen consumption without CO2 release) were not considered. In the table, a plus sign '+' indicates production and a minus sign '-' indicates consumption of metabolites. D The growth curve of *P. aeruginosa* colored in black. Additionally, for all phenotypes (P1, P3-P9) the growth curve is shown. To distinguish the different times when a certain phenotype did occur, an integrated boxplot is given below. E Comparison of predicted doubling times with experimental findings. Minimal medium and rich medium doubling times were shown.

got more metabolically diverse and a division of metabolic tasks and cooperation between phenotypes occurred. Coupled with decreased oxygen levels in the center, a fermenting

phenotype (P4) appeared and the produced acetate was consumed by phenotype P5, which appeared subsequently (Figure 3.4D). The core of the mature biofilm consisted mainly of acetate producers (P4) and metabolically inactive cells that had zero flux through the biomass reaction (P1). The next phase of biofilm formation was characterized by a highly dynamic cooperation and competition between phenotypes. Succinate was released, in addition to acetate, by a new phenotype P6. Fermenting phenotypes, P4 and P6, were most abundant and therefore quantities of acetate and succinate began to rise (Figure 3.4A and 3.4D). The newly available succinate was used again by the emerging phenotypes P7 and P8. The production and consumption of different amounts of acetate and succinate under varying oxygen conditions are due to difference in nutrient availability, as shown by independent FBA simulations (see S3 Text). Experimentally it has been shown that *P. aeruginosa* cultures are able to produce acetate, and succinate as fermentation products which also contribute to biofilm survival [50]. Additionally, it has been reported that *P. aeruginosa* is able to use succinate as a carbon source and that the addition of acetate or succinate increased the growth rate [161, 178]. In addition to experimental findings, our simulation identifies fermenting (P4, P6, P7) and absorbing phenotypes (P5, P7, P8) whose interactions contributes to community stability. After about 16.5 hours of simulation, only very small concentrations of oxygen remained and the mature biofilm could be divided into three layers: a metabolic inactive core and two fermenting outer layers (Figure 3.4A and 4B). Both fermenting layers consisted of acetate and succinate producers (P6, P9). First the outer fermenting layer was formed out of phenotype P6 which grew towards the edges in which the glucose concentration was still high. Afterwards the inner fermenting layer with phenotype P9 showed an additional fixation of CO2 by the anaplerotic pyruvate carboxylase reaction. In this context it is known that CO2 can exert both a positive or negative effect on growth of *Pseudomonas* [100, 65] and thus carbon fixation could be possible (more detailed discussion in S3 Text). Our simulation further suggests that CO2 fixation can have a positive effect on late phase biofilm survival.

Finally the inactive core increased in size and dominated the population after 20 hours with cell death and population decrease. We found oxygen to be the limiting factor (Figure 3.4B and 3.4D). Concerning anaerobic physiology, it has been reported that *P. aeruginosa* can grow in microaerobic and anoxic environments [169]. Anoxic growth has been shown either with nitrate or nitrite as alternative electron acceptors [24], or via arginine [206] and pyruvate

[50] fermentation by which the former allowed only minor growth and the latter supported survival only [169]. We tested the influence of nitrate as alternative electron acceptor in an additional simulation. When the population consisted mostly of metabolic inactive cells after 20 hours, $0.1mM$ nitrate was added. Shortly afterwards a new nitrate respiring phenotype P11 replaced the former dominant, metabolic inactive phenotype P1. Therefore, the almost dissolving biofilm culture could be reactivated by adding another terminal electron acceptor instead of oxygen (S3 Fig, [18]).

BacArena demonstrates how emergent metabolic phenotypes could contribute to community formation. We were able to make novel predictions on how these different phenotypes could contribute to biofilm integrity within a spatio-temporal context. Recently, a role of metabolic co-dependence between interior and peripheral cells for community stability, resilience, and antibiotic resistance has been described for *B. subtilis* biofilms [116]. Our simulation shows that a similar metabolic cooperation could be possible in *P. aeruginosa* biofilms between micro-aerobically fermenting and aerobic phenotypes. Novel treatments could try to first eliminate the protective outer layer and then target the metabolic cross-feeding of the inner layer to disrupt the overall biofilm structure, by targeting specific metabolic pathways particular to the corresponding phenotypes.

3.3.3 Integrated multi-species model of a human gut community

We used BacArena to model the multi-species community of the human gut (Figure 3.5). Since the human gut microbiota typically comprises 500-1000 species [47], we implemented a simplified human intestinal microbiota (SIHUMI) of seven species that has previously been characterized experimentally [11]. In a first condition (Figure 3.5A), we added all metabolites which can be consumed by at least one species to the environment except mucus glycans. *E. coli* dominated the community after the population reached a stable state at $16h$ (Figure 3.5B). This condition could correspond to a dysbiotic gut environment with intestinal bacterial overgrowth, in which *E. coli* dominates the human gut flora [19]. Interestingly, by adding a more realistic mucus glycan gradient to our model, we could revert the *E. coli* dominance (Figure 3.5D) and a spatial differentiation of the community between gut lumen and mucus layer emerged. The mucus layer was mostly dominated by *B. thetaiotaomicron*, which is

well known to degrade glycans [102]. This result is in accordance with experimental data, which showed the same niche separation between mucus degrading bacteria close to the gut epithelial layer and other microbes in the lumen [46]. Moreover, this spatial differentiation is indicative for a healthy gut microbiota since mucus degrading bacteria can occupy and defend the space close to the epithelium and consequently out-compete intruding pathogens [32]. An impaired mucus secretion can lead to inflammatory bowel disease, where the epithelial barrier is infiltrated by bacteria [189] which cause an inflammation of the gut wall. In this context, our results support recent evidence suggesting that metabolite secretion by the host may play a more important role in shaping the gut microbiome than the immune system itself [168]. Our model therefore predicts that metabolic gradients are relevant in shaping the gut community structure and ecology. This has some important implications in understanding the mucus barrier and indicates that dietary or metabolic treatments might be more relevant than immunosupressors in case of a disrupted mucosal microbiota.

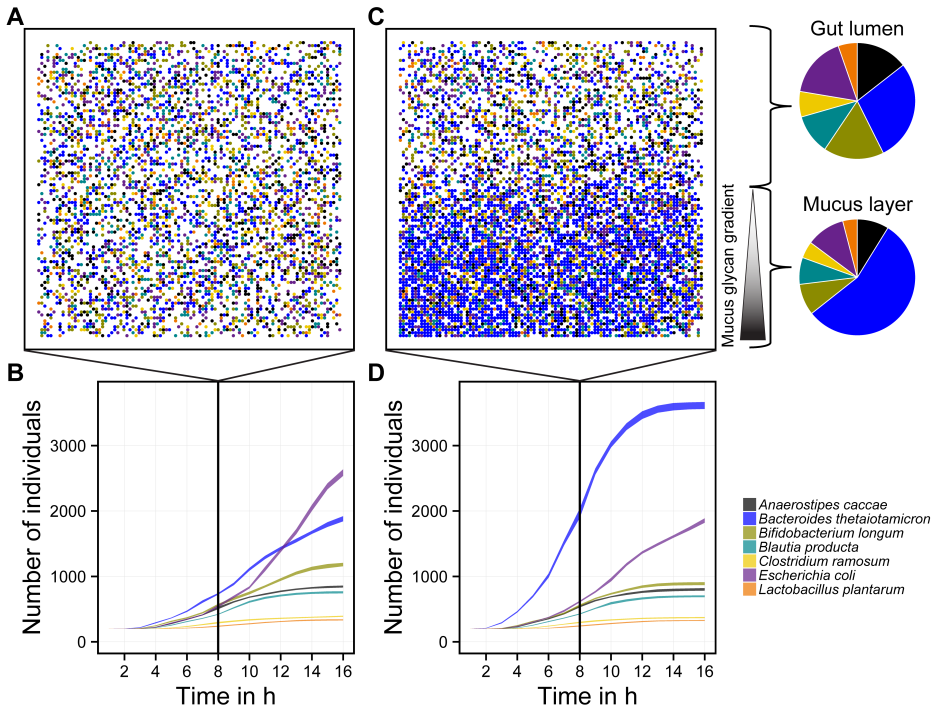

Figure 3.5: Multi-species community of a minimal human intestinal microbiota (SIHUMI) in rich medium. A Spatial population structure in the exponential phase after simulating 8 hours under a uniformly distributed rich medium (all possible metabolites that can be taken up are added to the environment), with B the growth curves of each species. C Spatial population structure in the exponential phase after 8 hours simulation time under a uniformly distributed rich medium with a spatial gradient of mucus glycans, with D the growth curves of each species. The curve range shows the standard deviation of 10 replicate simulations.

Next, we focused on the underlying metabolic mechanisms influencing the overall ecological structure of our setup which includes the mucus glycans (Figure 3.5C). As expected from human gut studies [56], we found the fermentation products succinate, acetate, lactate, propionate, and butyrate (Figure 3.6B) to be produced and, in some cases, exchanged between the microbes (Figure 3.6C). As for propionate, butyrate, and acetate, we could compare our predictions (Figure 3.6A) to the initial experimental study which describes the SIHUMI microbiota and in vitro co-culture experiments [11]. We found that the metabolite concentration ratios are comparable to experimental values with minimal higher butyrate and lower propionate concentrations (Figure 3.6A). Since BacArena allows to assess the metabolic phenotype of individual cells, we are able to derive hypotheses concerning cross-feeding of fermentation products. In particular, succinate was the metabolite with the most diverse metabolic exchange among the present metabolites (Figure 3.6C). This observation is in concordance with experimental findings suggesting an importance of succinate cross-feeding between human gut microbes [53]. In addition to succinate, acetate was also a key component to cross-feeding interactions between the microbes of our simplified community (Figure 3.6C). Acetate was produced by all species, except *B. longum* (Figure 3.6B). This might explain the experimentally observed high levels of acetate concentrations in the human large intestine [35], likely resulting from an over-production of acetate compared to its consumption. Furthermore, the relatively high concentration of acetate is also in concordance with experimental studies on the SIHUMI model microbiota (Figure 3.6A). As expected, lactate was mainly produced by the lactic acid bacteria *B. longum* and *L. plantarum*, and consumed by *B. producta*, *C. ramosum*, and *E.coli* (Figure 3.6B). Butyrate was released by *A. caccae* and *E.coli* and was not part of any cross-feeding interactions (Figure 3.6C). The remaining butyrate could therefore be potentially absorbed by the host epithelium as a main metabolite for energy conversion [12].

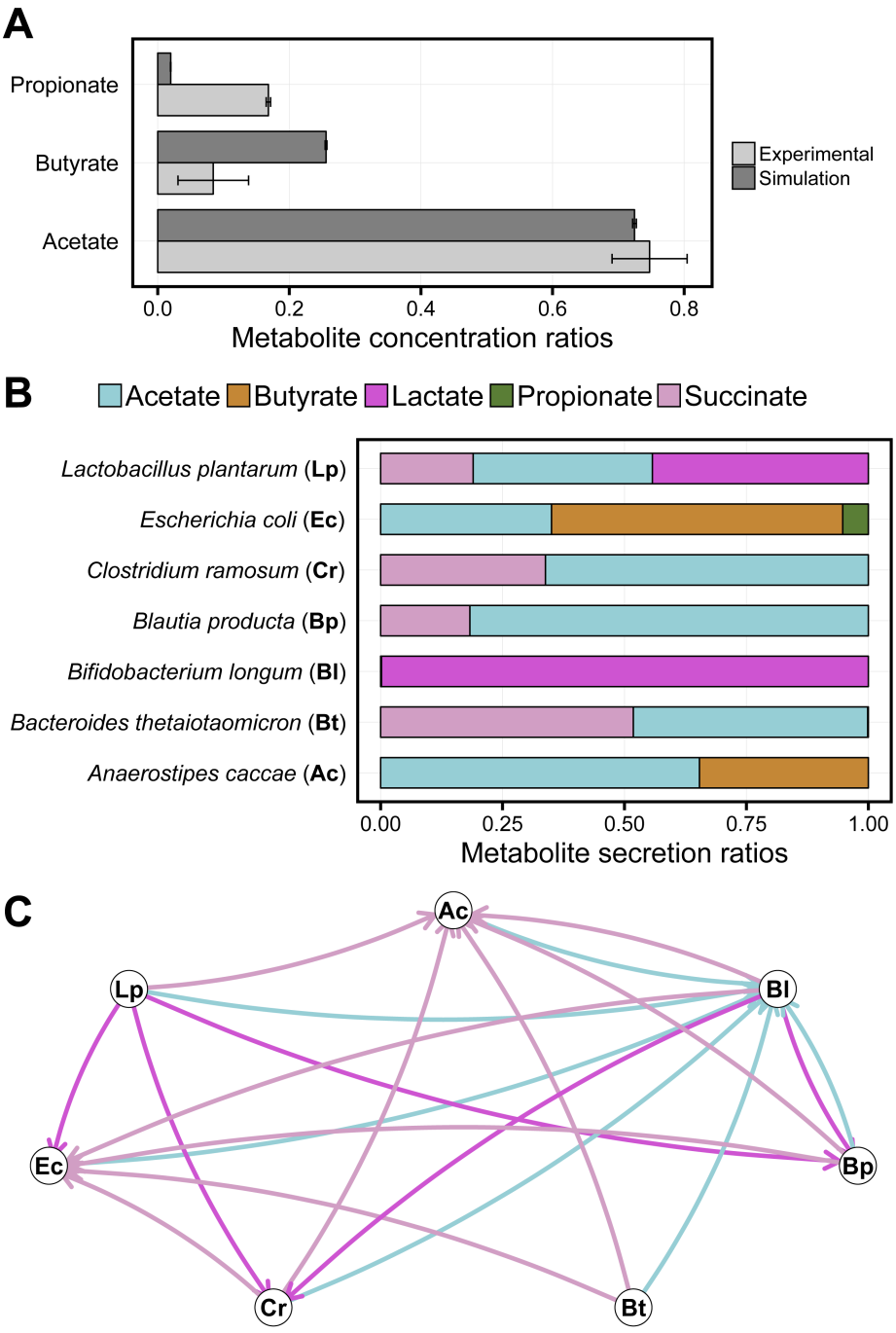

Figure 3.6: Influence of mucus glycan gradients on community dynamics. A Comparison of simulated metabolite concentrations with experimental values based on in vitro SIHUMI co-cultures [11]. B Metabolite secretion rates of different microbes in our SIHUMI model, determined by the overall metabolic secretion flux of the populations comprising all individuals. C Emerging metabolic interaction network of different fermentation products that can be exchanged between the microbe population in our SIHUMI model. Nodes represent species and edges represent exchanged metabolites, which are directed from the secreting species to the consuming species. The secretion and uptake was determined by the overall metabolic flux of the populations comprising all individuals.

To investigate the impact of alternative optimal FBA solutions on the reproducibility of our results, we randomized the selection of alternative optimal solutions and checked the simulations against each other (S4 Fig). We found that growth curves did not change with differing methods, which we expected since our simulated alternative optimal solutions have the same objective value (in our case the growth rate). Despite some metabolite concentrations variations (S4 Fig), the general trend was consistent and thus we concluded our results to be stable.

The predicted metabolite concentrations and cross-feeding interactions of our model (Figure 3.6C) give novel insights into how the simultaneous exchange of multiple fermentation products is relevant in shaping the human gut microbiota.

3.3.4 Conclusion

Following the systems biology paradigm, we presented a novel approach to study cellular communities. BacArena enables the analysis of interaction dynamics on the level of individuals and can therefore contribute to current efforts to move from correlative to functional explanations.

In context of a single-species biofilm of *P. aeruginosa*, we could show how a dynamic series of locally interacting metabolic phenotypes contributed to the emergence of an overall biofilm structure. We found that within species metabolic heterogeneity is an important contributor to community dynamics. The spatial differentiation in biofilms has been shown to have important implication in biofilm stability and integrity since the outer layer can act as protective barrier and the inner core can serve as a seed to initiate a new biofilm by supplying metabolites after antibiotic treatment [116, 185].

Additionally, we used BacArena to study the dynamics of gut microbes interacting within the epithelial mucus layer, which has important implications in inflammatory bowel disease [189]. As multi-scale modeling approaches become more relevant in studying the gut microbiome [90], BacArena provides an important contribution since it allows explore the relevance of metabolic interactions in the dynamics of such communities.

Chapter 4

From metagenomic data to personalized in silico microbiotas: Predicting dietary supplements for Crohn's disease

Manuscript in prepartion

Abstract

Crohn's disease (CD) is associated with an ecological imbalance of the intestinal microbiota, consisting of hundreds of species. The underlying complexity as well as individual differences between patients contributes to the difficulty to define a standardized treatment. Computational modeling can systematically investigate metabolic interactions between gut microbes to unravel novel mechanistic insights. In this study, we integrated metagenomic data of CD patients and healthy controls with genome-scale metabolic models into personalized in silico microbiotas. We predicted short chain fatty acid (SFCA) levels for patients and controls, which were overall congruent with experimental findings. As an emergent property, low concentrations of SCFA were predicted for CD patients and the SCFA signatures were unique to each patient. Consequently, we suggest personalized dietary treatments that could improve each patient's SCFA levels. The underlying modeling approach could aid clinical practice to find novel dietary treatment and guide recovery by rationally proposing food aliments.

4.1 Introduction

The human gut microbiota is composed of thousand different bacterial species with a large functional diversity that surpasses the human genome in terms of the collective gene pool [158]. Health promoting functions of the gut microbiota include the breakdown of otherwise indigestible dietary fibers and production of short chain fatty acids (SCFA) utilized by the human host [39].

Various human diseases, including inflammatory bowel disease (IBD), are associated with a loss of functional and taxonomic diversity of the gut microbiota [158]. The main symptom of IBD is inflammation of the gut epithelium [99]. Depending on the site of inflammation, IBD is distinguished in ulcerative colitis, which primarily affects the colon, and Crohn's disease (CD), in which different sites can be affected. Non-invasive treatments for CD include the intake of antibiotics [155] and steroid therapies, which suppress the immune system [203]. In addition, dietary change with a defined formula is used to ease the symptoms of the disease [210]. However, the success of these diet formulas varies between patients [73]. Additionally, after remission, patients have difficulties in finding an appropriate diet and often experience relapse. Considering the metabolic relevance of the human gut microbiota, it has been suggested that the dietary formula effects and reshapes the microbiota [93]. Overall, the microbial diversity is decreased in CD patients. A shortage of SCFAs [87] coincides with a decreased abundance of fermenting Firmicutes bacteria [124]. Microbial SCFAs have been recognized as important modulators of the immune system and as a nutrition source [75]. Butyrate, for example, is taken up as an additional energy source by the host [42], contributes to epithelial barrier integrity [152], and stimulates the immune system [59]. CD patients suffer from a low butyrate concentration [37], but its dietary supplementation can revert many of the IBD symptoms [164], highlighting the relevance of this particular SCFA in CD.

Given that the human gut microbiota is a complex microbial community with many different microbes that have varying metabolic potentials and substrate affinities [9], it becomes difficult to track the manifold ecological interactions that differ between CD patients and healthy individuals. Meta-omics approaches are generally used to characterize the microbiota and its metabolic potential [222]. However, these top-down approaches do not provide

mechanistic insights on the resilience of the microbiota and how perturbations, such as dietary treatments, may affect the system as a whole.

Bottom-up systems biology approaches can mechanistically describe biological systems and make biologically relevant predictions. In particular, constraint-based reconstruction and analysis (COBRA) has been successfully applied to model the metabolism of different species and make predictions on how perturbations affect the metabolic phenotype [145]. Briefly, genome-scale metabolic reconstructions are represented by the complete set of biochemical reactions derived from a genome annotation and organism-specific literature in a stoichiometric accurate manner [192]. Such high-quality manually-curated metabolic reconstructions are available for organisms from all three domains of life, such as *E. coli* [143], yeast [140], and human (e.g., [193]). Through the application of specific constraints (e.g., nutrient availability), the metabolic reconstructions can be converted into condition-specific models, which predict the reaction flux rates and growth yield under a given objective that is optimized using flux balance analysis (FBA) [145]. In a recent publication [10], we combined FBA with agent based modeling to simulate the ecology of microbial communities through the BacArena framework. Based on dietary metabolites in the environment, the metabolic models are constrained to simulate the metabolic states of each species, optimizing their growth yield. Metabolic interactions emerge from the exchange of metabolites between species and the environment. These interactions can influence the metabolite concentration and the microbial community by inducing cross-feeding or resource competition. Such COBRA-based approaches provide a powerful mean to investigate mechanistic links in complex biological systems, such as the human gut microbiota.

A recent study on pediatric CD sequenced the metagenomes of a North American cohort consisting of 26 healthy controls and 85 patients newly diagnosed with CD [110]. In their study, the authors could distinguish two clusters of patients: A cluster of 57 patients, which had a microbiota composition similar to the healthy controls, and a cluster of 28 patients that had a distinguished dysbiotic microbiota. Compared to controls, these dysbiotic patients had a strongly differing functional and microbial abundance profile.

Here, we retrieved the original metagenomic data of the 26 healthy controls and 28 dysbiotic patients [110] to simulate personalized in silico microbiotas with BacArena. We demonstrate that the simulated metabolic differences between patients and controls are con-

gruent with experimental findings. We further show that predicted individual specific SCFA signatures are unique to each patient. Based on these results, we then predict personalized dietary treatments that would improve the SCFA concentrations of each patient. With this work, we demonstrate the added value of performing computational modeling in conjunction with high-throughput data of individual microbiotas to predict mechanism-based personalized dietary intervention strategies for CD patients.

4.2 Methods

4.2.1 Retrieval of metagenomic data and pre-processing

Paired-end Illumina raw reads of a study on early onset Crohn's disease (CD) patients and healthy controls of a North American cohort [110] were retrieved from NCBI SRA under the accession: SRP057027. Based on the studies' definition of healthy and dysbiotic individual microbiotas [110], the samples were selected to a smaller subset of 26 healthy controls and 28 CD patients to capture the most pronounced differences in the individual microbial communities. The selected patients showed pronounced differences in their functional and microbial profile in the original study [110]. Furthermore, only the first measured time point was selected to represent newly diagnosed and yet untreated microbiotas. The reads were quality trimmed using Trimmomatic [17] with default parameters for paired-end Illumina sequences. To remove human contaminant sequences, the reads which were still paired after the quality control were mapped with default parameters using the software BWA [113] to the human genome version 38 (`http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/`).

4.2.2 Metagenomic mapping and abundance estimation

Using BWA [113], the pre-processed reads were mapped with default parameters onto a reference set of 773 genomes, which were selected according to a previous study [183] (see also below). Before mapping, the reference genomes of these organisms were combined into one file where each genome is represented as a chromosome. To filter out cross-mapped reads (reads mapped to multiple positions), samtools [114] was used to discard mapped reads

with a low-quality score. The coverage per genome (number of mapped reads normalized by genome size) was calculated using samtools. To reduce the number of false positives, we set a threshold of at least 1% genome coverage for each microbe in each human individual. In accordance to another pipeline [96], the resulting coverages were normalized for each individual to obtain the relative microbe abundances.

4.2.3 Microbial metabolic reconstructions

We retrieved published gut microbial metabolic reconstruction [183] from `http://vmh.life`. These microbes have been chosen according to their prevalence in the human gut and the availability of a genome sequences, and they have been extensively curated based on available physiological and biochemical data [183]. In average, a microbial reconstruction consisted of 933 +/- 139 metabolites, 1,198 +/- 241 reactions, and 771 +/- 262 genes.

4.2.4 Analysis of mapped abundance and reaction differences

The mapped microbial abundances for each individual were compared by computing the Bray-Curtis similarity and subsequent visualization with principal coordinate analysis (PCoA) using the R package *vegan* [41]. The unique reaction set of personalized in silico microbiota was determined by taking the union of all present microbe reactions. These reactions were retrieved from the corresponding metabolic models [183] of each microbe. PCoA was performed on the metabolic distance between each individual's reaction set similar to [183].

4.2.5 Setup, integration, and simulation of the personalized microbiota models

In the previous steps, the microbial abundance information for each individual was determined. The next step is to integrate this information into a personalized in silico microbiota for each person. Therefore, we used a previously established R package for community modeling [10], which represents bacteria as individuals in a grid environment that can exchange metabolites by secretion and uptake. Multiple species can be integrated into the environment with varying number of individuals per species. The dimensions of the two-dimensional

quadratic environment was set $0.025cm^2$ with 100 grid cells per side length. This resulted in 10,000 grid cells that could be potentially occupied by the microbes. To allow space for the in silico microbial community to grow, only 500 microbes were initially added to the grid environment. The relative microbial abundances were used to determine the number of microbes to be added per species (e.g., if one species has a relative abundance of 0.01, 5 microbes were added for this species on the entire grid). In case the calculated number of microbes resulted in decimal places, we rounded the final number to the next highest integer. All possible metabolites (union of metabolites that can be taken up by each microbe) were added to the environment with a minimal concentration of $0.2\mu M$ to provide a rich medium that is consistent between individuals. Therefore, metabolite concentrations that emerge from the simulations can be specifically attributed to the microbiota of each individual.

Once the in silico microbiota for each CD patient and healthy control have been setup in BacArena, the growth of each microbial model in the microbiota was sequentially for each time step. A total of 24 time steps were simulated, one per hour, corresponding to an overall simulation time of 24 hours. At each time step, the medium composition of each grid cell was updated as a function of the metabolites that were taken up or secreted by the occupying microbes. When a certain growth rate of a microbe occupying a grid cell was reached, a neighboring free grid cell could be occupied by the microbe. If no neighboring grid cell was available, then cells do not duplicate. To reduce the complexity of the model, we simulated a well-mixed environment in which metabolite concentrations are uniformly distributed and microbes move randomly.

The R package *sybil* [62] was used for constraint based modeling. ILOG CPLEX was used as a linear programming solver. The computations were carried out on high performance computer clusters.

4.2.6 Analysis of simulation results

After the simulation, each personalized in silico microbiota was primarily analyzed in terms of the microbe abundance and metabolite concentrations. Since the simulations include temporal dynamics with different time points, we chose the last time point (24h) for our analysis and comparison between individuals. This allowed the in silico microbial communities

enough time to consume and produce metabolites, and to reach a steady state. The microbial abundances were determined by assessing the number of microbes in each personalized in silico microbiota. The vector of microbial abundances was then compared by computing the Bray-Curtis similarity and subsequently visualized with PCoA. Abundances of specific taxa were calculated by summing up the relative abundances of each corresponding representative. The abundances of the most differing taxa were tested for significant differences between healthy controls and CD patients with the Wilcoxon rank-sum test [208].

Metabolite concentrations were determined by their molar concentration in the environment at the end of the simulation (t=24h). The concentration of the most relevant metabolites, butyrate, propionate, isobutyrate, L-lactate, and acetate, were assessed and tested for significant differences between the personalized in silico microbiotas of healthy controls and of CD patients using the Wilcoxon rank-sum test. To investigate the influence of each microbial taxa on the metabolite concentrations, we further evaluated the metabolic fluxes of each microbe in the personalized in silico microbiota. For each taxa, the reaction fluxes in all corresponding microbes were summed up.

4.2.7 Definition of personalized dietary treatments

After identifying the metabolic signatures influencing CD and the corresponding microbes causing these differences between healthy controls and CD patients, we predicted metabolites that could revert these differences. Therefore, we analyzed each genome-scale microbial metabolic model separately.

According to their presence in each personalized in silico microbiota, the set of microbes was selectively analyzed for every individual. Each personalized in silico microbiota was then simulated in a rich medium containing all possible metabolite with flux uptake constraints of $1 mmol gDW^{-1} h^{-1}$ and the biomass as well as the production of SCFAs (butyrate, propionate, isobutyrate, L-lactate, acetate) were optimized for separately. To enhance the growth of beneficial bacteria, we selected metabolites based on the ability of the CD low abundant microbes (e.g., Clostridia, Bacteroides) to uptake these nutrients over the CD high abundant microbes (e.g., Gammaproteobacteria, Bacilli). We then added the selected metabolites iteratively to the in silico medium with a maximal flux uptake constraint of $1000 mmol gDW^{-1} h^{-1}$ to

investigate whether the fermentation products increased or decreased. Based on these simulations, the added metabolites which had a positive effect (recovering metabolite production to healthy levels) were then collected and used as the personalized dietary treatment for each individual.

We tested the effect of the treatment on the personalized in silico microbiota of CD patients by adding a 100 times higher concentration of the predicted treatment metabolites to the in silico rich diet containing $0.2\mu M$ for each metabolite. The personalized in silico microbiota simulations and analyses were then carried out as described above.

4.2.8 Quantification and statistical analysis

Differences between healthy controls and CD patients were assessed with the Wilcoxon rank-sum test implemented in R. For the healthy controls our group size was 26 individuals and for the CD group 28 individuals.

4.2.9 Data and software availability

The scripts to construct and simulate the individual specific microbiota models as well as the analysis scripts are available on GitHub: `https://github.com/euba/CodeBase`.

4.3 Results

The aim of the present study was to predict in silico novel personalized dietary treatments for CD and investigate individual differences. We simulated personalized in silico microbiotas consisting of hundreds microbial metabolic models as defined by published metagenomic data of healthy controls and CD patients [110] using a novel computational modeling approach [10] in which we combined FBA with agent based modeling to simulate the ecology of microbial communities through the BacArena framework. Based on dietary metabolites in the environment, the metabolic models are constrained to simulate the metabolic states of each species, optimizing their growth yield. Metabolic interactions emerge from the exchange of metabolites between species and the environment. These interactions can be used to gain further insight into metabolic differences that may contribute to CD and to propose

modeling-assisted dietary intervention strategies for CD patients (Figure 4.1). We describe differences between healthy controls and CD patients based on SCFAs as well as microbial abundances, which we validated with existing experimental knowledge. Individual differences within patients and controls were assessed to find the SCFA signature specific to each individual microbiota. Based on the individual microbiotas, personalized dietary treatments, such as supplementation of pectin and different glycans, were predicted to equilibrate the SCFA concentrations and promote healthier SCFA concentrations. Taken together, our work demonstrates the use of computational modeling to integrate existing high-throughput data of individual microbiotas and mechanistically predict novel personalized dietary treatments for CD.

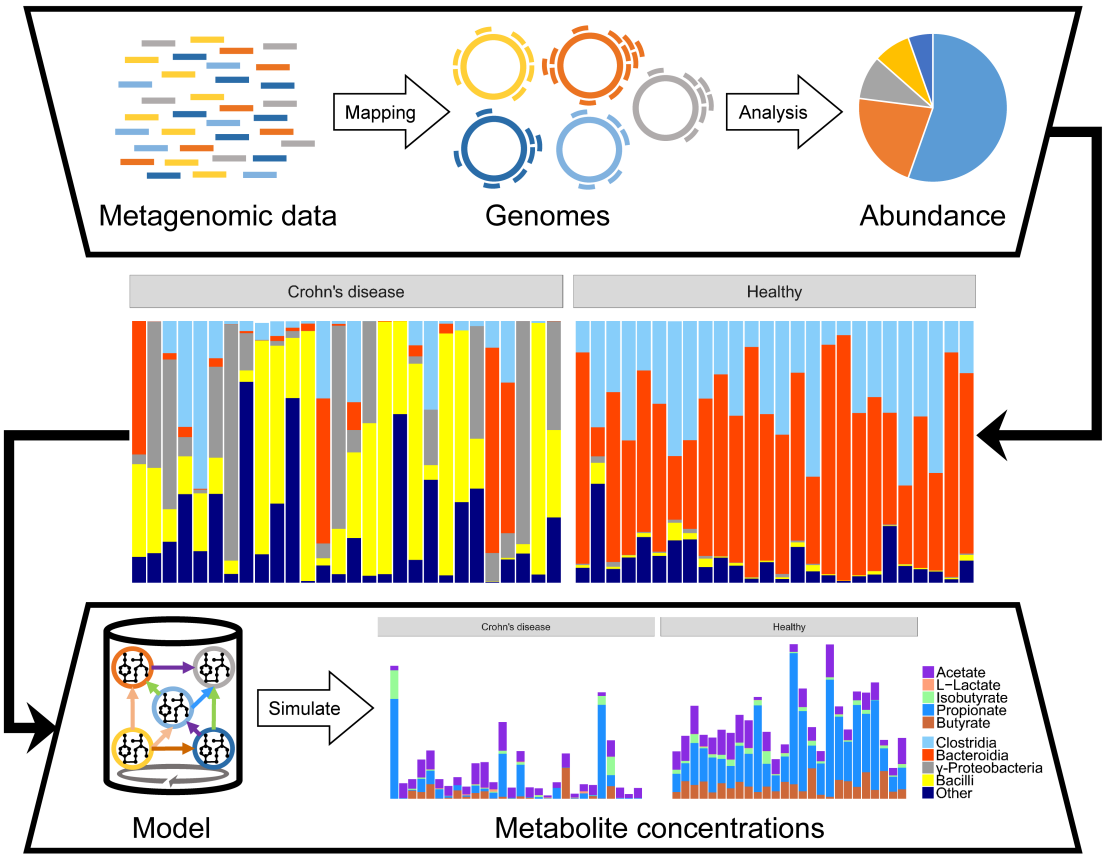

Figure 4.1: Computational framework used to create personalized metabolic models of gut microbial communities. Published metagenomic data were integrated into an in silico microbiota model for each CD patient and healthy control to simulate emergent metabolite concentrations.

4.3.1 Microbial differences between healthy controls and CD patients

We ensured that our computational workflow (Figure 4.1) would not alter the reported microbial differences between healthy controls and of dysbiotic CD patients [110]. The workflow mapped the published metagenomic data of healthy controls and CD patients onto the genome sequences of the 773 gut microbial strains, for which metabolic reconstructions were available [183]. In average, 283 +/- 240 of the 773 microbial strains were covered in the personalized in silico microbiota (Supplementary Figure C.1). Notably, the smallest in silico microbiota contained only eight microbes, while the biggest had 713 of the 773 microbial strains. There were seven out of 54 in silico microbiotas that had less than 40 of the 773 microbes. While CD patients had generally less microbes, there were also some healthy controls with less than 40 microbes and some CD patients with more than 600 microbes (Supplementary Figure C.1). Overall, the personalized in silico microbiota captured 73.5 +/- 16% of the relative microbial abundance from the original metagenomic reads. We could observe a clear separation of the healthy controls and CD patients when assessing the microbial differences based on microbial abundances captured by the in silico microbiota (Figure 4.2A), which was independent on the used similarity metrics (Supplementary Figure C.2). The most pronounced differences between the healthy and the CD individuals were due to significantly higher abundance of Bacilli and Gammaproteobacteria ($p<0.05$, Wilcoxon rank-sum test) and significantly lower abundance of Bacteroidia and Clostridia ($p<0.001$, Wilcoxon rank-sum test) in CD patients (Figure 4.2D).

We then simulated the personalized in silico microbiota, which were inoculated with 500 microbes on a grid with 10,000 cells for 24 hours in the BacArena framework and analyzed whether the microbial abundances changed compared to the initial (metagenomic data driven) abundances. At the end of the simulation, the grid was populated by an average of 5902 +/- 1743 microbes corresponding to an average grid occupation of 59 +/- 17%. Overall, the simulated abundances recapitulate the initial microbial differences, demonstrating that the in silico microbiotas were stable over time in BacArena (Figure 4.2B). However, the abundance ratios of four out of 28 genera were consistently higher in CD patients based on the simulated abundance, but lower based on the mapped data (Figure 4.3A). In contrast, the mapped abundance data showed good agreement with the abundances reported in the original

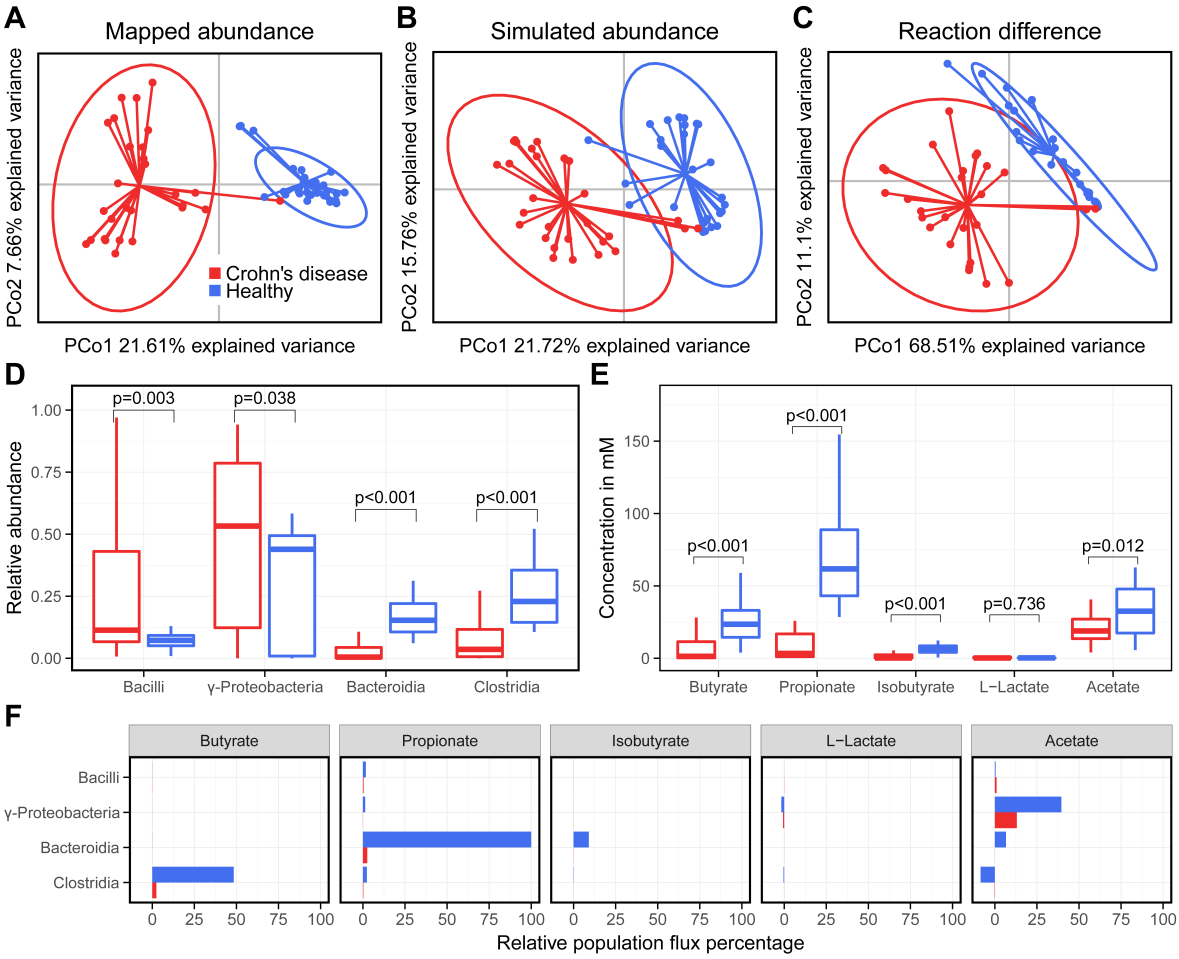

Figure 4.2: Metabolic and microbial group variability between healthy controls and Crohn's disease patients. Similarities were assessed based on a principle coordinate analysis (PCoA) of the reaction content with Jaccard distance (A), mapped abundance with Bray Curtis dissimilarity (B), and simulated abundances with Bray Curtis dissimilarity (C). Based on the simulation, relative abundances (D) and metabolite concentrations of fermentation products (E) were compared (p-value determined by Wilcoxon rank-sum test). Microbial metabolic activities were displayed as the total population flux (F).

study (Figure 4.3A, Supplementary Figure C.3). This discrepancy can be explained by the CD patients having a lower diversity of microbes, which led to a higher predicted abundance for the present genera.

Taken together, our workflow recapitulates the reported microbial differences between controls and CD patients [110]. Furthermore, the simulation results of the personalized in silico microbiota in BacArena illustrate that these microbes can also co-exist, in a similar relative abundance, as stable microbial communities in silico.

4.3.2 Emergent metabolic differences between healthy controls and CD patients

We investigated whether the difference in microbial abundance in the personalized in silico microbiota also corresponded to differences in metabolic repertoires. That is: Are there microbes in some of the personalized in silico microbiota that are unique and may help to distinguish healthy and CD microbiota? In average, each personalized in silico microbiota consisted of 3,332,957 +/- 285,848 metabolic reactions belonging to 3,036 +/- 424 unique metabolic reactions. The presence and absence pattern of the unique reactions in the in silico microbiota varied between individuals as well as between the two groups (Figure 4.2C). Interestingly, based on the reaction content, the first two principal components explained almost 80% of the variation in the data (Figure 4.2C), and were mainly driven by the presence of transport reactions for fibers (Supplementary Table C.1). The observed reaction based separation is consistent with the aforementioned differences in microbial classes (Figure 4.2D) and the distinct fiber metabolizing properties of Bacteroides.

SCFAs are important energy precursors and interact with the human immune system [59]. We analyzed the secretion of SCFAs after 24 hours by each personalized in silico microbiota, consisting of a myriad microbial models that can individually consume and secrete the SCFAs, to establish whether known microbiota-level differences in SCFA production could be reproduced by our computational modeling approach. The SCFAs butyrate, propionate, isobutyrate, and acetate were significantly lower in CD patients ($p<0.05$, Wilcoxon rank-sum test, Figure 4.2E). Only L-lactate levels were slightly higher in CD patients. To check for the validity of the simulated metabolite concentrations, we compared our results with an independent experimental study [86]. The qualitative difference between CD patients and healthy controls were consistent with our simulations (Figure 4.3B). However, the predicted concentrations of butyrate and propionate were three times higher in controls than in CD patients (Figure 4.3B), which is much higher than the reported difference, probably due to the absence of the host cells in our model setup that can take up butyrate and propionate produced by the microbiota [38]. Overall, our results confirm that the personalized in silico microbiotas also recapitulate known differences in SCFA production levels in healthy and CD individuals. An advantage of using computational modeling is that we can determine which

microbes, or microbial classes, in the in silico microbiota caused the predicted differences in SCFA production. Therefore, we analyzed the summed uptake and secretion fluxes of each microbial class in the personalized in silico microbiota. We found that Clostridia were responsible for the production of 50% of the total butyrate, Bacteroidia produced almost 100% of the total propionate and about 10% of the total isobutyrate, Bacilli produced small quantities (<5% of the total concentration) of L-lactate, and Gammaproteobacteria produced almost 50% of the total acetate (Figure 4.2F). Notably, in healthy controls, acetate was taken up by Clostridia illustrating a cross-feeding mechanism between Gammaproteobacteria and Clostridia. These results demonstrated how changes in representatives of the main microbial classes can result in differences in SCFA production capabilities that differ significantly between healthy controls and CD patients.

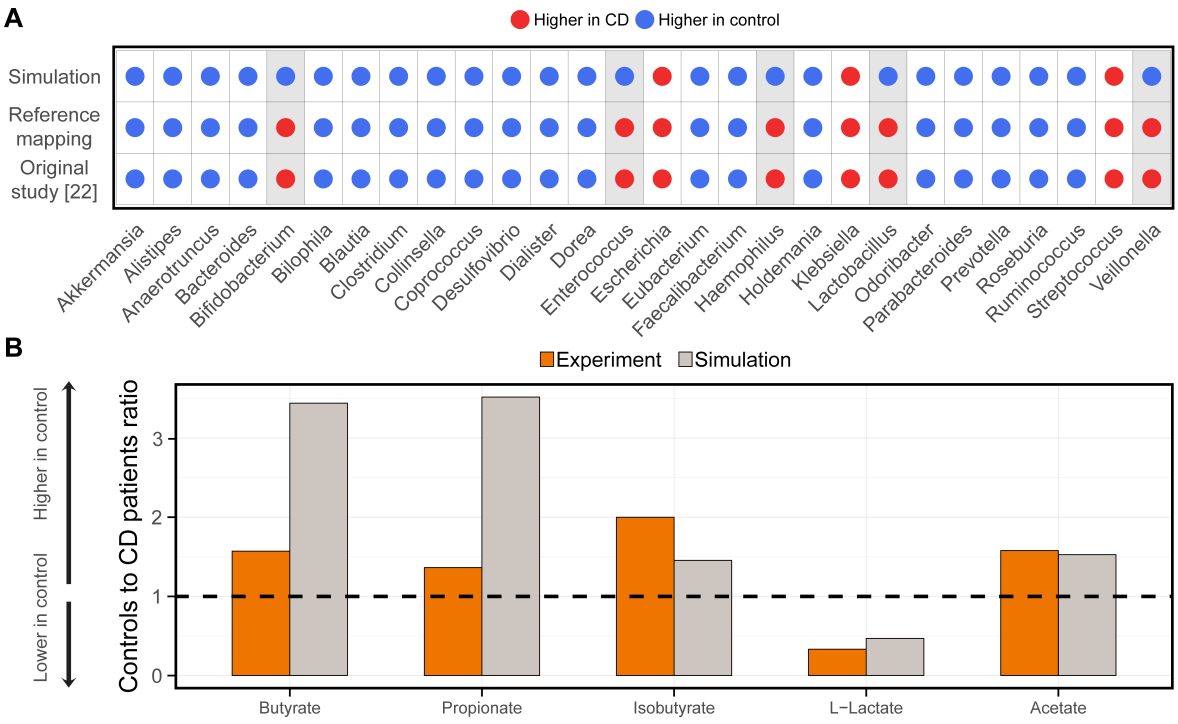

Figure 4.3: Qualitative comparison of simulation results with experimental values. Experimental relative abundances of microbial genera (A) were retrieved from the original study [110] and compared with the abundances based on the mapped reads and simulations (t=24h). (B) Metabolite concentrations were retrieved from an independent experimental study [86] and compared with the simulations (t=24h) based on the mean concentration ratios of healthy controls and CD patients.

4.3.3 SCFA production profiles are patient-specific

The original metagenomic study [110] reported the most distinct microbial differences between the healthy controls and the CD patients but also variability within the groups. Accordingly, the simulated relative microbial abundance also varied between the individuals (Figure 4.4, left). We next investigated how much the predicted SFCA production capability varied between CD patients. Two (CD10, CD11) out of 28 CD patients had butyrate levels that were comparable to the mean of healthy controls (mean concentration of 7.5 and 25.8mM for CD and controls respectively). This could be explained by the higher metabolic activity of Clostridia species in these patients (Figure 4.4, right). In three cases (CD2, CD4, CD22), the concentration of isobutyrate was higher in CD patients (Figure 4.4) compared to the mean concentration in healthy controls (mean concentration of 4.9 and 7.1mM for CD and controls respectively). Two of these patients (CD2, CD22) also had propionate levels comparable to the mean of healthy controls (mean concentration of 25 and 87.9mM for CD and controls respectively), which is congruent with the high metabolic activity of isobutyrate and propionate producing Bacteroides species (Figure 4.4, right). Twelve out of the 28 patients showed increased L-lactate concentrations (mean concentration of 0.7 and 0.3mM for CD and controls respectively), which can be attributed to the metabolic activity of Bacilli and other taxa (Figure 4.4). This is also congruent with the observation, that CD patients had an overall higher L-lactate concentration (Figure 4.2E). Five patients (CD11, CD16, CD17, CD19, and CD25) showed acetate levels that were comparable to the mean of healthy controls (mean concentration of 21.1 and 32.2mM for CD and controls respectively). This can be mostly attributed to the metabolic activity of Bacilli and Gammaproteobacteria (Figure 4.4, right). Overall, these results indicated that every patient has a specific signature of SCFAs, in which some metabolites have levels comparable to healthy controls. This observation can be explained by the metabolic activity of the present microbiota, indicating that metabolic stimulation of the native CD microbiota may be able to revert some of the differences between CD and healthy controls.

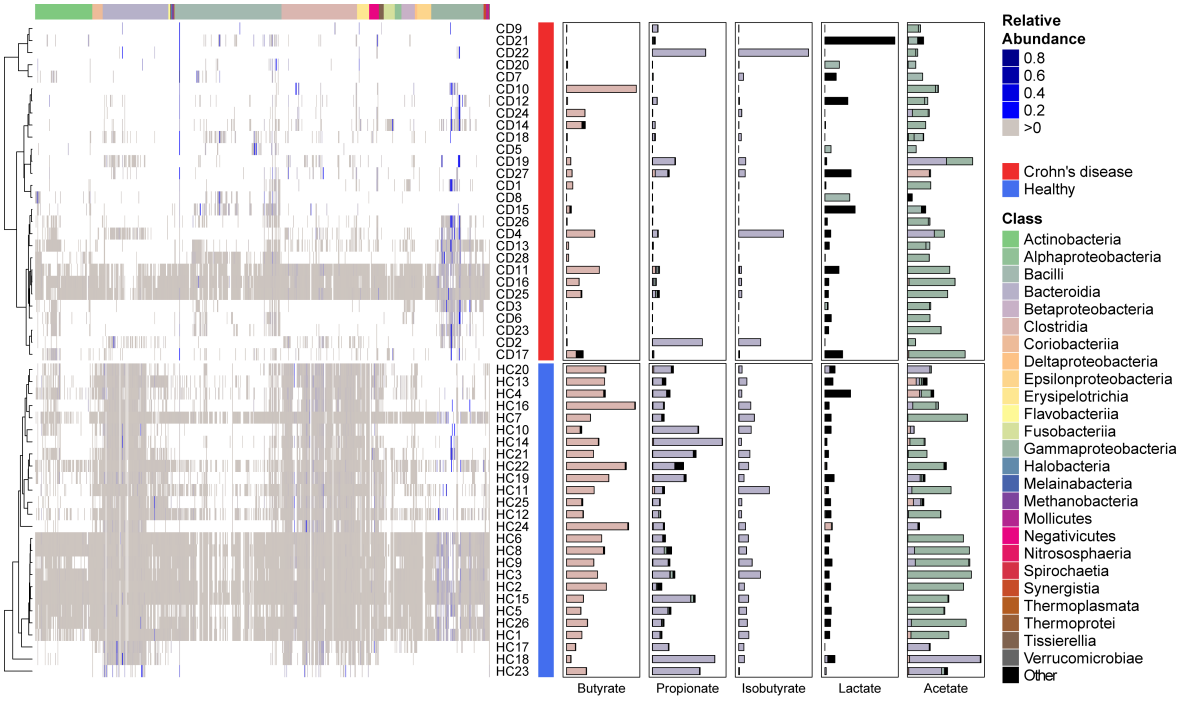

Figure 4.4: Individual variability between CD patients and healthy controls. The presence of different microbes is indicated by a gray color and the relative abundance by a blue color scale. Microbial taxa are based on the class level. Predicted metabolite concentrations are based on simulations. The microbial contribution to the concentrations are based on metabolic fluxes.

4.3.4 Personalized dietary intervention strategies to normalize SCFA production capabilities of the personalized in silico microbiota

Defined dietary regimes are one possible treatment strategy for CD patients [210]. However, the success of this treatment varies between CD patients [38]. As butyrate is often low in CD patients and its dietary supplementation has been reported to ease gastrointestinal symptoms of CD patients [164], we investigated whether we could design personalized dietary interventions that would restore the SCFA production to levels commonly reported in healthy individuals. We approached this problem by predicting first whether increasing each dietary compound, present in the in silico rich diet, could individually lead to a more healthy level of each of the five SCFAs in any microbial model present in a given CD patient in silico microbiota (Figure 4.5A). Interestingly, the number of the predicted dietary metabolites to be supplemented was highly specific for each patient and ranged between 1 and 55 metabolites (median of 19 metabolites) (Figure 4.5B). For four out of the 28 CD patients, our described prediction approach did not identify any metabolites that could normalize the production of

any SCFA. These four patients had a higher abundance of Gammaproteobacteria and Bacilli, while major SCFA producers, belonging mostly to Bacteroidetes and Clostridia were largely absent. For the remaining 24 CD patients, the most prominent category of the predicted metabolites were mucus glycans and glycosaminoglycans (Figure 4.5B). In particular, pectin supplementation was predicted to be a good dietary metabolite for 17 out of 24 CD patients (Supplementary Figure C.4). Other prevalent metabolites included various specific human produced mucus glycans and hepan/hyaluronan proteoglycan degradation products as well as plant-derived larch arabinogalactan, lavanbiose, and amylose.

We then added all of these identified metabolites to each of the personalized in silico microbiota to ensure that the community could also produce healthier SCFA levels. Each personalized in silico microbiota was simulated for 24 hours in the personalized supplemented diets. The success of the in silico dietary interventions varied between the patients (Figure 4.5C). Overall, the most successful individual metabolite level restoration was obtained for butyrate, propionate, and acetate, whereas the in silico treatment was less successful for isobutyrate and L-lactate (Figure 4.5C). Overall, the SCFA concentration differences between CD patients and healthy controls (Figure 4.5D) were also improved with respect to butyrate, propionate, and acetate, whereas isobutyrate and L-lactate did not show improved levels compared to the untreated concentrations (Figure 4.5D). The in silico treatments had only small effects on the relative species abundances in the personalized in silico microbiotas as they were not able to restore a more healthy ratio of microbial species (Figure 4.5E) due to the dysbiotic patients lacking the relevant microbes found in healthy individuals. Therefore, our results showed quantitatively improved levels of SCFAs on the individual patient level as well as on the differences between patients and healthy controls. To re-establish a healthier microbiota composition, we would need to account for food-borne microbes or probiotics in addition to the dietary treatment.

4.4 Discussion

We created personalized in silico microbiota of healthy controls and CD patients by integrating metagenomic data into a bottom-up systems biology framework (Figure 4.1). Recent approaches have successfully integrated metagenomic data to model the ecological dynamics

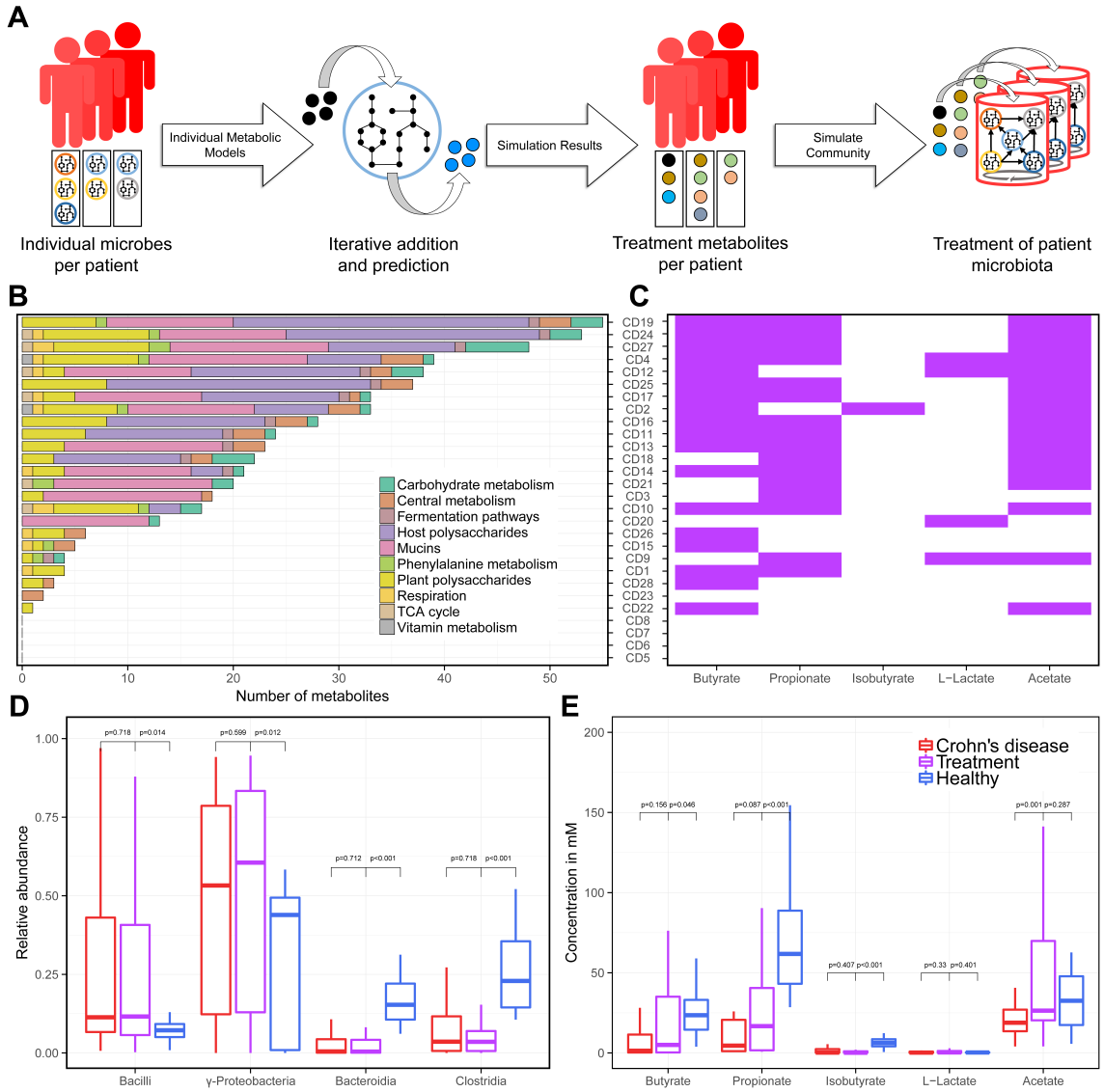

Figure 4.5: Individual treatment prediction for each CD patient. For the prediction of treatment metabolites (A), single metabolic models of microbes for each patient were optimized for the production of the target metabolites with iterative dietary additions. Panel (B) shows broader categories of the predicted metabolites and (C) shows the response (metabolite increase of 25%) of each patient in purple. Panel (D) and (E) show the relative abundance and metabolite concentrations.

of the human gut microbiota [8] but lack the metabolic aspect, which plays an important role for human health and disease [197]. Therefore, the added benefit of our modeling approach is that we combine metabolism and ecology to investigate the metabolic activity of the human gut microbiota.

To find strong differences between CD patients and healthy controls, we selected data of dysbiotic patients that were defined by their microbial distance to healthy controls [110].

Expectedly, we could reproduce the microbial differences originally reported in the study (Figure 4.2A). Moreover, our reference based assessment was consistent with the reference independent analysis in the original study (Figure 4.3A), which further demonstrates that the set of 773 AGORA microbes capture the most common human gut microbes [183]. When comparing the abundance of specific genera (Figure 4.3A), the community simulations predict differing ratios for four out of 28 genera, which indicates a minor variability in the simulations that did not affect the overall differences (Figure 4.2B). The main microbial differences between CD patients and healthy controls can be attributed to a decreased abundance of Bacteroidia and Clostridia as well as an increased abundance of Bacilli and Gammaproteobacteria in CD patients (Figure 4.2D), which was in accordance with an independent experimental study [92] and characteristic for a dysbiotic microbiota. Our results therefore capture strongly dysbiotic CD microbiotas, which is, however, only one specific case of CD [110]. This approach allows us to address fundamental questions in CD dysbiosis and to mechanistically describe how the microbiota can shape metabolite concentrations, which is less understood so far.

The simulated SCFA concentrations represent emergent properties of our models that could not be achieved by the metagenomic data alone. Therefore, we could simulate clinical relevant metabolite concentrations, known to be differentially regulated in CD [86]. Interestingly, we could see differences in SCFA levels between healthy controls and CD patients (Figure 4.2E). In particular, higher concentrations of acetate, propionate, butyrate, and isobutyrate as well as a lower concentration of L-Lactate in controls (Figure 4.2E). We could also validate these qualitative differences with experimental literature [86] (Figure 4.3B). Based on the quantitative ratios between controls and patients, butyrate and propionate were higher in our simulations than in experiments (Figure 4.3B). This apparent discrepancy could be explained by the uptake of butyrate and propionate by the host [38], which we did not include and is therefore a limitation of our current modeling set up. SCFAs, in general, have been associated with healthy gut functions, such as energy conversion of the host as well as immune stimulation [75]. Butyrate, in particular, mediates the immune system [59] and influences the tight junctions between epithelial cells [152]. Moreover, butyrate, as well as propionate, are carbon sources for colonocytes [162, 29]. Taken together, the added value of our modeling approach is that we can predict these qualitative changes in SCFA levels, which

we can attribute to specific microbial metabolic activity.

We identified which microbes are responsible for the production of the SCFA (Figure 4.2F). Clostridia produced mainly butyrate, explaining its lower concentration in CD patients (Figure 4.2E), who had generally lower Clostridia abundances (Figure 4.2D). The Clostridia *Faecalibacterium* and *Roseburia* are known to be the main butyrate producers [120]. Interestingly, these two Clostridia were decreased in abundance in CD patients (Figure 4.3B). We identified novel metabolic interaction patterns, such as the consumption of acetate by Clostridia (Figure 4.2F). *In vitro* experiments have demonstrated cross-feeding interactions between Clostridia and Bifidobacterium species [12]. Based on our simulation, we suggest that these interactions take also place in an ecologically complex microbiota and that they are of high relevance in the human gut. These metabolic interactions link microbes with metabolites and demonstrate that we capture in silico the gut microbiota as a whole.

Our personalized in silico microbiota modeling approach permitted the investigation of individual differences between CD patients and healthy controls (Figure 4.4). Overall, we found that healthy controls have a higher diversity of microbes than CD patients, which is also confirmed by experimental knowledge [124]. Consequently, controls have more comparable SCFA levels (Figure 4.4), indicating metabolic consistency through functional redundancy of microbes [88]. Interestingly, some patients had some SCFA but not all levels comparable with the controls, which could be attributed to a higher metabolic activity of health relevant microbes in those patients (Figure 4.4). One could speculate that the microbiota of CD patients can compensate some metabolic differences but lacked functional redundancy and diversity to consistently establish a healthy SCFA signature (Figure 4.4). This observation further underlines the importance of a diverse microbiota, which can complement potential metabolic shortcomings between microbes and consistently produce SCFAs. Further studies could investigate the importance of keystone species in this context, which have a low abundance but high metabolic activity and thus ecological relevance [198].

In our in silico treatment predictions, we take the individual factors into account by designing dietary supplements that would compensate the individual differences (Figure 4.5A). Most of the predicted treatment metabolites were mucus glycans, glycosaminoglycans, and plant polysaccharides (Figure 4.5B), further indicating that fibers are relevant in shaping the gut microbiota metabolism [127, 102]. Particularly, pectin was predicted as a potential

treatment for the majority of patients, which further points towards the dietary relevance of this compound [127]. Plant fibers and host glycans can influence the gut microbiota by stimulating Clostridia and Bacteroidia species [55], which produce butyrate and propionate, respectively (Figure 4.2F). Interestingly, the predicted metabolite cocktails were different for each patient (Figure 4.5B, Supplementary Figure C.4). In clinical practice, a standard dietary formula in form of exclusive enteral nutrition is used to treat patients with CD [210]. However, not every patient responds equally well to different diet formulations, which vary in their fiber content [115]. Current knowledge is limited when defining personalized diets because of the complexity of the human gut microbiota and its intricate response to different diets. Some patients suffer from relapse when switching to a normal diet after successful remission [13]. In such cases, our modeling-based predictions could give novel directions on aliments based on a patient's microbiota. Finally, a dietary treatment strategy for an individual is not likely to be static. Using computational modeling in conjunction with metagenomic data, the dietary treatment could be readily redefined and adjusted to match the patient's need. To our knowledge, such modeling-guided dietary treatment approach is not available yet for Crohn's disease patients. As a next step, our predictions need to be validated in a nutritional trial. Then, our systematic approach to defining personalized nutrition therapies could guide clinicians and nutritionists in designing new, personalized diet-based treatments.

Testing our in silico dietary treatments on each patient's microbiota, we found an improvement in SCFA levels. Butyrate, propionate, and acetate showed an overall success in shifting levels, while isobutyrate and L-lactate were less successful (Figure 4.5C, 4.5E), since those SCFAs only had a minor difference between controls and patients (Figure 4.2E). The overall microbe abundance did also not shift significantly in the treatment condition (Figure 4.5D), because patients had a lower diversity from the start (Figure 4.4) and could not acquire the necessary microbes to compensate their abundance profile. Further studies could simulate the effect of adding specific microbe models as a treatment, which could be integrated in our framework. Furthermore, human metabolism could be integrated with the in silico microbiota to investigate the reciprocal effect on the host, and, for instance, the effect of colorectal cancer cells that might be affected by butyrate concentrations [172].

Several studies emphasize the need for computational models to discover novel hypothesis and mechanisms for microbiota associated diseases [14, 191, 90]. Our approach introduces

metabolism as an additional emergent property of the microbiota yielding into novel mechanistic insight of SCFA production by microbial communities. So far, our approach ignores for simplicity the spatial component, which could have important implications in shaping the ecology and possible metabolic interactions. An extension for possible treatment strategies includes the simulation of probiotics and fecal transplantation. In fact, our model could be used as an additional workflow for donor optimization of fecal transplantation [148] by finding the most appropriate microbiota. Most importantly, the computational modeling approach that we presented here is not limited to the application of CD but can be applied to any metagenomic data set and disease with microbial dysbiosis. Taken together, we present a powerful, expandable, versatile computational modeling approach that permits to yield novel insight into metabolic interactions emerging from personalized metagenomic data and to predict personalized dietary intervention strategies.

Acknowledgements

We want to thank Dr. Almut Heinken for classifying the treatment metabolites and giving useful comments on the analysis of the results. We also want to thank Mr. Marouen Ben Guebilla, Mr. Alberto Noronha, and Mr. Federico Baldini for giving useful comments on the manuscript. This work was supported by an ATTRACT program grant (FNR/A12/01), and an Aides a la Formation-Recherche (FNR/6783162) grant.

Chapter 5

Concluding remarks

The data avalanche of high throughput technologies guided the discovery of an incredibly complex microbial ecosystem within the human body, the gut microbiota. These analyses characterized thousands of different microbial species that have a high functional potential, which collectively surpasses the genetic component of the human host [158]. However, several challenges came with the analysis of high throughput technologies, which need to be addressed with novel methodologies and concepts. One of the biggest challenges is the question of correlation and causality: While multiple diseases such as inflammatory bowel disease and obesity have been associated with an altered or dysbiotic gut microbiota [30], it is not clear if this is merely a consequence or the direct cause of the disease. Systems biology approaches can contribute to the field of human gut microbiota research by providing these mechanistic insights in form of hypothesis and predictions that can drive further experimental and clinical research.

Methods in systems biology are usually based on networks that can be topologically analyzed. In constraint based reconstruction and analysis (COBRA), such networks are based on the metabolism (reactions that are connected by their transformed metabolites) of an organisms that is deduced from the corresponding genome annotations of enzymes with databases containing previous biochemical knowledge. Simulations can be performed on these networks to find fluxes in metabolic pathways under different environmental conditions. This can aid in the understanding of the functional potential of the human gut microbiota and how different conditions affect this complex microbial community.

The work of this thesis describes the use and development of COBRA approaches to un-

derstand the human gut microbiota. First, metabolic differences between human gut microbes were assessed in terms of automatically derived metabolic reconstructions. This analysis revealed that the metabolism can be more variable between species than their phylogenetic relationships, which underlines the relevance of functional and microbial diversity in the human gut microbiota. In the second study, a modeling approach, BacArena, was developed, which simulates the metabolic interactions between microbes in a community. This novel approach was applied to a single species biofilm model and a simplified seven species community representative of the human gut. In the third study, the modeling approach was used to simulate more realistic human gut microbiotas consisting of up to hundreds of microbes by integrating metagenomic data of Crohn's disease patients as well as healthy controls. In this applied analysis, experimentally known metabolite concentrations were recapitulated, as well as species abundances, and novel personalized treatments could be predicted that shift the key metabolite concentrations of the CD patients to values that are more similar to healthy controls. Taken together, the work of this thesis demonstrates the relevance of systems biology approaches to analyze the human gut microbiota ecology and associated diseases.

5.1 Investigating topological similarities of metabolic networks from human gut microbes

The human gut microbiota is a complex ecological network of thousands of different microbes with varying functions. Studies based on high throughput data driven analyses have characterized the metabolic potential of the human gut microbiota [158] and also attributed these functions to different microbial groups [222]. However, reference based assessments are scarce and often investigate only a specific group of functions.

To investigate the functions of a representative set of human gut microbes, genome scale metabolic reconstructions were automatically reconstructed based on the available genome sequence of each microbe (Chapter 2). In this analysis, metabolic reactions involved in lipid and short chain fatty acid pathways were found to be most distinguishing between microbes, which is consistent with traditional taxonomic methods that classify microbes

[70]. Interestingly, the metabolic variability between microbes is more sensitive towards strain specific differences than the phylogenetic relationships. Conversely, phylogenetic differences when comparing higher taxonomic degrees are more sensitive than metabolic differences. This indicates that there is a functional redundancy in the higher taxonomic levels and a relatively high functional diversity in the lower taxonomic degrees such as strains. This is thus consistent with human gut studies that show functional redundancy [88] as well as studies that show strain specific metabolic differences [135]. Furthermore, this might explain why the human gut microbiota consists of so many different species that are able to coexist with each other by occupying specific niches in which only specific tasks are required, e.g., the utilization of specific carbohydrates. In addition, this also highlights the need to take multiple microbes into account to capture the metabolic capacity of each strain and ultimately the ecological complexity of the human gut microbiota.

5.2 Modeling the metabolic ecology of intestinal microbial communities

Since the human gut microbiota is an ecosystem of different microbes interacting with each other, it is important that modeling approaches take this complexity into consideration. As discussed in Chapter 1 such methods need to be scalable to represent as many microbes as possible. Recent developments in constraint based reconstruction and analysis (COBRA) have established a variety of community modeling paradigms that aim to simulate the metabolic interactions between microbes (Chapter 1).

Chapter 3 describes an approach, BacArena, in which the interactions between microbes is modelled by combining COBRA with individual based modeling, a method classically used for ecological studies [91]. In BacArena, microbes of different species are represented as individuals in a discrete spatial environment, in which they can secrete and uptake metabolites in distinct time steps. The metabolites can diffuse through this environment, so that the products of one microbe can reach the neighbouring fellow microbes. Therefore, complex metabolic interactions can emerge, in which individuals of the same or different species can engage in metabolic cross-feeding. Intra-species cross-feeding can induce metabolic differ-

entiation of a single species population in which resources can be utilized more efficiently. In Chapter 3, this has been applied and discussed in the context of pathogens to establish and maintain a resilient biofilm structure.

Further in Chapter 3, BacArena was then applied to model inter-species cross-feeding interactions of intestinal bacteria. In a simplified community model of seven gut associated bacteria, concentration gradients of mucus glycans have been identified as important in shaping the spatial community structure and formation of niches in which microbes can co-exist. This is also in congruence with experimental studies, which show a differential microbe distribution in the mucus layer compared to the lumen. Furthermore, this further strengthens the hypothesis that metabolic resources are important for shaping the intestinal microbial community [168]. With the investigation of cross-feeding interactions, novel hypothesis were created with respect to the exchange of fermentation products. While these analyses were focussed and calibrated on a simplified community of seven species, BacArena was designed to scale up to efficiently model more realistic gut microbiotas consisting of several hundred species.

5.3 Personalized patient simulations and guiding potential treatments

Recent efforts in gut microbiota research have tried to model the ecological dynamics of different human individuals to understand the underlying mechanisms that shape this complex ecosystem [8]. However, no study has yet taken up the challenge to model patient specific ecological dynamics of the microbiota metabolism. Chapter 4 applied a simplified version of BacArena (Chapter 3) to simulate the dynamics and differences of Crohn's disease patients compared to healthy controls. Crohn's disease is an inflammatory bowel disease that is characterized by a lower diversity of gut microbes [158].

The patient specific models were simulated based on a well mixed environment in which individuals move randomly and metabolites are uniformly distributed. This therefore represents a condition that is likely to be present in the gut lumen. For estimating patient specific abundances or population sizes of different species, we mapped the metagenomic reads of

each patient to the genomes of a representative set of 773 human gut microbes. The mapped reads were then normalized between patients to get the relative abundances of microbes which could then be integrated in our community model by scaling the number of individuals per species to represent the abundance proportions. A standard rich diet was then applied for each patient and simulated the metabolic potential of the personalized microbiotas.

The simulations reproduced experimentally known metabolic differences between Crohn's disease patients and healthy controls. Furthermore, the mechanism by which these differences occur was predicted in terms of production and uptake by specific microbes. This knowledge was then used to design personalized diets for each patient to achieve metabolite levels that are more similar to healthy controls. Such predictions could guide potential clinical treatments to have a rational designed diet that could lead to a better patient recovery.

5.4 Current challenges and future perspectives

Personalized human gut microbiota models of the metabolism can be used to generate novel mechanistic insights into metabolic interactions within this complex ecosystem. Predictions can be helpful for microbiota associated diseases of which the mechanism of current treatments such as fecal microbiota transplantation, standardized diets, or probiotics is poorly understood. Further experimental validation is required to demonstrate the usefulness of the predictions made in this thesis. Moreover, future studies can take this work as a basis to design experiments in a systematic manner. It would be also interesting to integrate the human metabolism into the microbiota models to account for host-symbiont relationships and metabolic exchange.

When choosing the appropriate modeling paradigm for a specific scientific question, it is important to consider all assumptions and limitations of the underlying method to interpret the simulation correctly. Depending on the question, preference should be taken to the most simple and predictive model to reduce the simulation complexity. Simplified representations of the human gut consisting of <10 species, such as the one discussed in Chapter 3, can be modeled with more complicated set-ups including temporal and spatial dynamics, whereas comprehensive simulations with more microbes benefit from further abstraction. Moreover, to ultimately answer ecological questions related to the human gut microbiota it is important

to take as much microbes into account as possible. Such questions include: Why is the microbiota so diverse? Why are some diseases associated with a less diverse microbiota? Can diseases be treated by increasing the metabolic and microbial diversity?

To study host-microbe interactions, it would be interesting to study other systems that are easier to manipulate, such as insect symbioses. In this context, modeling studies with COBRA have been conducted based on the pea aphid symbiont *Buchnera aphidicola* [195, 214]. Mostly, these studies focussed on the metabolic reduction and adaptation of the symbiont to its host. However, no metabolic modeling has been performed yet on more complex intestinal communities of insect hosts, which could prove useful for gnotobiotic insect models that can experimentally manipulated to carry a defined set of microbes as their gut microbiota [190].

Further studies need to develop more thorough data mining approaches to analyze the results of complex microbiota models. In addition, community models need to be scalable and therefore need more efficient algorithms. Given the large amount of high-throughput sequencing patient data it would be preferential to use a method that is suitable for data integration to create personalized models. By integrating these top-down and buttom-up approaches, predictive models can be designed that give novel mechanistic insights and new directions in the field of gut microbiota research.

Bibliography

[1] Adler, J. and Dahl, M. M. (1967). A method for measuring the motility of bacteria and for comparing random and non-random motility. *Microbiology*, 46(2):161–173.

[2] Adler, P., Bolten, C. J., Dohnt, K., Hansen, C. E., and Wittmann, C. (2013). Core fluxome and metafluxome of lactic acid bacteria under simulated cocoa pulp fermentation conditions. *Applied and Environmental Microbiology*, 79(18):5670–5681.

[3] Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature biotechnology*, 31(6):533–538.

[4] Amir, E. D., Davis, K. L., Tadmor, M. D., Simonds, E. F., Levine, J. H., Bendall, S. C., Shenfeld, D. K., Krishnaswamy, S., Nolan, G. P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature biotechnology*, 31(6):545–552.

[5] An, D., Na, C., Bielawski, J., Hannun, Y. A., and Kasper, D. L. (2011). Membrane sphingolipids as essential molecular signals for Bacteroides survival in the intestine. *Proceedings of the National Academy of Sciences*, 108(Supplement 1):4666–4671.

[6] Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., de Vos, W. M., Brunak, S., Dore, J., MetaHIT Consortium (additional members), Weissenbach, J., Ehrlich, S. D., and Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, 473(7346):174.

[7] Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST server: rapid annotations using subsystems technology. *BMC genomics*, 9(1):75.

[8] Bashan, A., Gibson, T. E., Friedman, J., Carey, V. J., Weiss, S. T., Hohmann, E. L., and Liu, Y.-Y. (2016). Universality of human microbial dynamics. *Nature*, 534(7606):259–262.

[9] Bauer, E., Laczny, C. C., Magnusdottir, S., Wilmes, P., and Thiele, I. (2015). Phenotypic differentiation of gastrointestinal microbes is reflected in their encoded metabolic repertoires. *Microbiome*, 3(1):55.

[10] Bauer, E., Zimmermann, J., Baldini, F., Thiele, I., and Kaleta, C. (2017). BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. *PLOS Computational Biology*, 13(5):e1005544.

[11] Becker, N., Kunath, J., Loh, G., and Blaut, M. (2011). Human intestinal microbiota: characterization of a simplified and stable gnotobiotic rat model. *Gut Microbes*, 2(1):25–33.

[12] Belenguer, A., Duncan, S. H., Calder, A. G., Holtrop, G., Louis, P., Lobley, G. E., and Flint, H. J. (2006). Two routes of metabolic cross-feeding between *Bifidobacterium adolescentis* and butyrate-producing anaerobes from the human gut. *Applied and Environmental Microbiology*, 72(5):3593–3599.

[13] Belluzzi, A., Brignola, C., Campieri, M., Pera, A., Boschi, S., and Miglioli, M. (1996). Effect of an enteric-coated fish-oil preparation on relapses in Crohn's disease. *New England Journal of Medicine*, 334(24):1557–1560.

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

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

[16] Blaut, M. and Clavel, T. (2007). Metabolic diversity of the intestinal microbiota: implications for health and disease. *The Journal of nutrition*, 137(3):751S–755S.

[17] Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

[18] Borriello, G., Werner, E., Roe, F., Kim, A. M., Ehrlich, G. D., and Stewart, P. S. (2004). Oxygen limitation contributes to antibiotic tolerance of *Pseudomonas aeruginosa* in biofilms. *Antimicrobial agents and chemotherapy*, 48(7):2659–2664.

[19] Bouhnik, Y., Alain, S., Attar, A., Flour, B., Raskine, L., Sanson-Le Pors, M. J., and Rambaud, J.-C. (1999). Bacterial populations contaminating the upper gut in patients with small intestinal bacterial overgrowth syndrome. *The American journal of gastroenterology*, 94(5):1327–1331.

[20] Braendle, C., Miura, T., Bickel, R., Shingleton, A. W., Kambhampati, S., and Stern, D. L. (2003). Developmental origin and evolution of bacteriocytes in the aphid–*Buchnera* symbiosis. *PLoS biology*, 1(1):e21.

[21] Bush, K. (2012). Antimicrobial agents targeting bacterial cell walls and cell membranes. *Revue scientifique et technique (International Office of Epizootics)*, 31(1):43–56.

[22] Cani, P. D., Lecourt, E., Dewulf, E. M., Sohet, F. M., Pachikian, B. D., Naslain, D., De Backer, F., Neyrinck, A. M., and Delzenne, N. M. (2009). Gut microbiota fermentation of prebiotics increases satietogenic and incretin gut peptide production with consequences for appetite sensation and glucose response after a meal. *The American journal of clinical nutrition*, 90(5):1236–1243.

[23] Cardinale, B. J., Palmer, M. A., Swan, C. M., Brooks, S., and Poff, N. L. (2002). The influence of substrate heterogeneity on biofilm metabolism in a stream ecosystem. *Ecology*, 83(2):412–422.

[24] Carlson, C. A. and Ingraham, J. L. (1983). Comparison of denitrification by *Pseudomonas stutzeri*, *Pseudomonas aeruginosa*, and *Paracoccus denitrificans*. *Applied and Environmental Microbiology*, 45(4):1247–1253.

[25] Carr, R. and Borenstein, E. (2014). Comparative analysis of functional metagenomic annotation and the mappability of short reads. *PLoS One*, 9(8):e105776.

[26] Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., Kaipa, P., Karthikeyan, A. S., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Pujar, A., Shearer, A. G., Travers, M., Weerasinghe, D., Zhang, P., and Karp, P. D. (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research*, 38(suppl_1):D473–D479.

[27] Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S. Y., Shearer, A. G., Tissier, C., Walk, T. C., Zhang, P., and Karp, P. D. (2008). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research*, 36(suppl_1):D623–D631.

[28] Chiu, H.-C., Levy, R., and Borenstein, E. (2014). Emergent biosynthetic capacity in simple microbial communities. *PLoS computational biology*, 10(7):e1003695.

[29] Clausen, M. R. and Mortensen, P. B. (1995). Kinetic studies on colonocyte metabolism of short chain fatty acids and glucose in ulcerative colitis. *Gut*, 37(5):684–689.

[30] Clemente, J. C., Ursell, L. K., Parfrey, L. W., and Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. *Cell*, 148(6):1258–1270.

[31] CLP (2017). Coin or clp. `http://projects.coin-or.org/Clp/`.

[32] Collado, M. C., Derrien, M., Isolauri, E., de Vos, W. M., and Salminen, S. (2007). Intestinal integrity and *Akkermansia muciniphila*, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Applied and Environmental Microbiology*, 73(23):7767–7770.

[33] Collins, M., Lawson, P., Willems, A., Cordoba, J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., and Farrow, J. (1994). The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *International Journal of Systematic and Evolutionary Microbiology*, 44(4):812–826.

[34] Coyte, K. Z., Schluter, J., and Foster, K. R. (2015). The ecology of the microbiome: networks, competition, and stability. *Science*, 350(6261):663–666.

[35] Cummings, J., Pomare, E., Branch, W., Naylor, C., and Macfarlane, G. (1987). Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut*, 28(10):1221–1227.

[36] David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., and Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505(7484):559–563.

[37] De Preter, V., Joossens, M., Ballet, V., Shkedy, Z., Rutgeerts, P., Vermeire, S., and Verbeke, K. (2013). Metabolic profiling of the impact of oligofructose-enriched inulin in Crohn's disease patients: a double-blinded randomized controlled trial. *Clinical and translational gastroenterology*, 4(1):e30.

[38] Den Besten, G., Lange, K., Havinga, R., van Dijk, T. H., Gerding, A., van Eunen, K., Müller, M., Groen, A. K., Hooiveld, G. J., Bakker, B. M., and Reijngoud, D.-J. (2013a). Gut-derived short-chain fatty acids are vividly assimilated into host carbohydrates and lipids. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 305(12):G900–G910.

[39] Den Besten, G., van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D.-J., and Bakker, B. M. (2013b). The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *Journal of lipid research*, 54(9):2325–2340.

[40] Dillon, R. and Dillon, V. (2004). The gut bacteria of insects: nonpathogenic interactions. *Annual Reviews in Entomology*, 49(1):71–92.

[41] Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, 14(6):927–930.

[42] Donohoe, D. R., Garge, N., Zhang, X., Sun, W., O'Connell, T. M., Bunger, M. K., and Bultman, S. J. (2011). The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell metabolism*, 13(5):517–526.

[43] Douglas, A. E. (2011). Lessons from studying insect symbioses. *Cell host & microbe*, 10(4):359–367.

[44] D'Souza, G., Waschina, S., Pande, S., Bohl, K., Kaleta, C., and Kost, C. (2014). Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. *Evolution*, 68(9):2559–2570.

[45] Dupont, C. L., Rusch, D. B., Yooseph, S., Lombardo, M.-J., Richter, R. A., Valas, R., Novotny, M., Yee-Greenbaum, J., Selengut, J. D., Haft, D. H., Halpern, A. L., Lasken, R. S., Nealson, K., Friedman, R., and Venter, J. C. (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *The ISME journal*, 6(6):1186.

[46] Earle, K. A., Billings, G., Sigal, M., Lichtman, J. S., Hansson, G. C., Elias, J. E., Amieva, M. R., Huang, K. C., and Sonnenburg, J. L. (2015). Quantitative imaging of gut microbiota spatial organization. *Cell host & microbe*, 18(4):478–488.

[47] Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., and Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science*, 308(5728):1635–1638.

[48] Eddelbuettel, D., François, R., Allaire, J., Chambers, J., Bates, D., and Ushey, K. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

[49] Edwards, J. and Palsson, B. Ø. (2000). The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97(10):5528–5533.

[50] Eschbach, M., Schreiber, K., Trunk, K., Buer, J., Jahn, D., and Schobert, M. (2004). Long-term anaerobic survival of the opportunistic pathogen *Pseudomonas aeruginosa* via pyruvate fermentation. *Journal of bacteriology*, 186(14):4596–4604.

[51] Evaldson, G., Heimdahl, A., Kager, L., and Nord, C. (1982). The normal human anaerobic microflora. *Scand J Infect Dis Suppl*, 35(1):9–15.

[52] Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., and Huttenhower, C. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS computational biology*, 8(7):e1002606.

[53] Ferreyra, J. A., Wu, K. J., Hryckowian, A. J., Bouley, D. M., Weimer, B. C., and Sonnenburg, J. L. (2014). Gut microbiota-produced succinate promotes *C. difficile* infection after antibiotic treatment or motility disturbance. *Cell host & microbe*, 16(6):770–777.

[54] Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., and Punta, M. (2013). Pfam: the protein families database. *Nucleic acids research*, 42(D1):D222–D230.

[55] Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R., and White, B. A. (2008). Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature reviews. Microbiology*, 6(2):121.

[56] Flint, H. J., Duncan, S. H., Scott, K. P., and Louis, P. (2007). Interactions and competition within the microbial community of the human colon: links between diet and health. *Environmental microbiology*, 9(5):1101–1111.

[57] Flint, H. J., Duncan, S. H., Scott, K. P., and Louis, P. (2015). Links between diet, gut microbiota composition and gut metabolism. *Proceedings of the Nutrition Society*, 74(1):13–22.

[58] Frank, D. N., Amand, A. L. S., Feldman, R. A., Boedeker, E. C., Harpaz, N., and Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences*, 104(34):13780–13785.

[59] Furusawa, Y., Obata, Y., Fukuda, S., Endo, T. A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T., Takahashi, M., Fukuda, N. N., Murakami, S., Miyauchi, E., Hino, S., Atarashi, K., Onawa, S., Fujimura, Y., Lockett, T., Clarke, J. M., Topping, D. L., Tomita, M., Hori, S., Ohara, O., Morita, T., Koseki, H., Kikuchi, J., Honda, K., Hase, K., and Ohno, H. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory t cells. *Nature*, 504(7480):446.

[60] Gareau, M. G., Sherman, P. M., and Walker, W. A. (2010). Probiotics and the gut microbiota in intestinal health and disease. *Nature Reviews Gastroenterology and Hepatology*, 7(9):503–514.

[61] Garrity, G. M., Bell, J. A., and Lilburn, T. G. (2004). Taxonomic outline of the prokaryotes. Bergey's manual of systematic bacteriology. *Springer, New York, Berlin, Heidelberg*.

[62] Gelius-Dietrich, G., Desouki, A. A., Fritzemeier, C. J., and Lercher, M. J. (2013). sybil–efficient constraint-based modelling in R. *BMC systems biology*, 7(1):125.

[63] Germerodt, S., Bohl, K., Lück, A., Pande, S., Schröter, A., Kaleta, C., Schuster, S., and Kost, C. (2016). Pervasive selection for cooperative cross-feeding in bacterial communities. *PLoS computational biology*, 12(6):e1004986.

[64] Gibbons, R. and Kapsimalis, B. (1967). Estimates of the overall rate of growth of the intestinal microflora of hamsters, guinea pigs, and mice. *Journal of bacteriology*, 93(1):510.

[65] Gill, C. O. and Tan, K. H. (1979). Effect of carbon dioxide on growth of *Pseudomonas fluorescens*. *Applied and Environmental Microbiology*, 38(2):237–240.

[66] GLPK (2017). GNU Linear Programming Kit. *http://www.gnu.org/software/glpk/*.

[67] Gonnerman, M. C., Benedict, M. N., Feist, A. M., Metcalf, W. W., and Price, N. D. (2013). Genomically and biochemically accurate metabolic reconstruction of *Methanosarcina barkeri* Fusaro, iMG746. *Biotechnology journal*, 8(9):1070–1079.

[68] Gorochowski, T. E., Matyjaszkiewicz, A., Todd, T., Oak, N., Kowalska, K., Reid, S., Tsaneva-Atanasova, K. T., Savery, N. J., Grierson, C. S., and Di Bernardo, M. (2012). BSim: an agent-based tool for modeling bacterial populations in systems and synthetic biology. *PloS one*, 7(8):e42790.

[69] Gosset, G. (2005). Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate: sugar phosphotransferase system. *Microbial Cell Factories*, 4(1):14.

[70] Gower, J. C. and Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of classification*, 3(1):5–48.

[71] Granger, B. R., Chang, Y.-C., Wang, Y., DeLisi, C., Segrè, D., and Hu, Z. (2016). Visualization of metabolic interaction networks in microbial communities using VisANT 5.0. *PLoS computational biology*, 12(4):e1004875.

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

[73] Griffiths, A. M., Ohlsson, A., Sherman, P. M., and Sutherland, L. R. (1995). Meta-analysis of enteral nutrition as a primary treatment of active Crohn's disease. *Gastroenterology*, 108(4):1056–1067.

[74] Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., and DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science*, 310(5750):987–991.

[75] Guarner, F. and Malagelada, J.-R. (2003). Gut flora in health and disease. *The Lancet*, 361(9356):512–519.

[76] Guckert, J. B., Ringelberg, D. B., White, D. C., Hanson, R. S., and Bratina, B. J. (1991). Membrane fatty acids as phenotypic markers in the polyphasic taxonomy of methylotrophs within the Proteobacteria. *Microbiology*, 137(11):2631–2641.

[77] Gupta, R. S. (2011). Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. *Antonie Van Leeuwenhoek*, 100(2):171–182.

[78] Gurobi (2017). Gurobi. `http://www.gurobi.com`.

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

[80] Heinken, A., Sahoo, S., Fleming, R. M. T., and Thiele, I. (2013). Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut microbes*, 4(1):28–40.

[81] Heinken, A. and Thiele, I. (2015a). Anoxic conditions promote species-specific mutualism between gut microbes in silico. *Applied and Environmental Microbiology*, 81(12):4049–4061.

[82] Heinken, A. and Thiele, I. (2015b). Systematic prediction of health-relevant human-microbial co-metabolism through a computational framework. *Gut Microbes*, 6(2):120–130.

[83] Heirendt, L., Thiele, I., and Fleming, R. M. T. (2017). DistributedFBA.jl: high-level, high-performance flux balance analysis in julia. *Bioinformatics*, 33(9):1421–1423.

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

[85] Holland, J. H. (1992). Complex adaptive systems. *Daedalus*, pages 17–30.

[86] Hove, H. and Mortensen, P. B. (1995). Influence of intestinal inflammation (IBD) and small and large bowel length on fecal short-chain fatty acids and lactate. *Digestive diseases and sciences*, 40(6):1372–1380.

[87] Huda-Faujan, N., Abdulamir, A., Fatimah, A., Anas, O. M., Shuhaimi, M., Yazid, A., and Loong, Y. (2010). The impact of the level of the intestinal short chain fatty acids in inflammatory bowel disease patients versus healthy subjects. *The open biochemistry journal*, 4:53.

[88] Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207.

[89] IBM (2017). IBM ILOG CPLEX. *http://www.ibm.com/us-en/marketplace/ibm-ilog-cplex*.

[90] Ji, B. and Nielsen, J. (2015). From next-generation sequencing to systematic modeling of the gut microbiome. *Frontiers in genetics*, 6.

[91] Judson, O. P. (1994). The rise of the individual-based model in ecology. *Trends in Ecology & Evolution*, 9(1):9–14.

[92] Kaakoush, N. O., Day, A. S., Huinao, K. D., Leach, S. T., Lemberg, D. A., Dowd, S. E., and Mitchell, H. M. (2012). Microbial dysbiosis in pediatric patients with Crohn's disease. *Journal of clinical microbiology*, 50(10):3258–3266.

[93] Kaakoush, N. O., Day, A. S., Leach, S. T., Lemberg, D. A., Nielsen, S., and Mitchell, H. M. (2015). Effect of exclusive enteral nutrition on the microbiota of children with newly diagnosed Crohn's disease. *Clinical and translational gastroenterology*, 6(1):e71.

[94] Kandler, O. (1983). Carbohydrate metabolism in lactic acid bacteria. *Antonie van Leeuwenhoek*, 49(3):209–224.

[95] Kanehisa, M. and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30.

[96] Karlsson, F. H., Nookaew, I., and Nielsen, J. (2014). Metagenomic data utilization and analysis (MEDUSA) and construction of a global gut microbial gene catalogue. *PLoS computational biology*, 10(7):e1003706.

[97] Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L., and Gordon, J. I. (2011). Human nutrition, the gut microbiome, and immune system: envisioning the future. *Nature*, 474(7351):327.

[98] Kaul, H. and Ventikos, Y. (2013). Investigating biocomplexity through the agent-based paradigm. *Briefings in bioinformatics*, 16(1):137–152.

[99] Khor, B., Gardet, A., and Xavier, R. J. (2011). Genetics and pathogenesis of inflammatory bowel disease. *Nature*, 474(7351):307.

[100] King, A. D. and Nagel, C. W. (1975). Influence of carbon dioxide upon the metabolism of *Pseudomonas aeruginosa*. *Journal of Food Science*, 40(2):362–366.

[101] Klitgord, N. and Segrè, D. (2010). Environments that induce synthetic microbial ecosystems. *PLoS computational biology*, 6(11):e1001002.

[102] Koropatkin, N. M., Cameron, E. A., and Martens, E. C. (2012). How glycan metabolism shapes the human gut microbiota. *Nature reviews. Microbiology*, 10(5):323.

[103] Kumar, V. S., Dasika, M. S., and Maranas, C. D. (2007). Optimization based automated curation of metabolic reconstructions. *BMC bioinformatics*, 8(1):212.

[104] LaBauve, A. E. and Wargo, M. J. (2012). Growth and laboratory maintenance of *Pseudomonas aeruginosa*. *Current protocols in microbiology*, pages 6E–1.

[105] Laczny, C. C., Pinel, N., Vlassis, N., and Wilmes, P. (2014). Alignment-free visualization of metagenomic data by nonlinear dimension reduction. *Scientific reports*, 4.

[106] Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Grarup, N., Jørgensen, T., Brandslund, I., Nielsen, H. B., Juncker, A. S., Bertalan, M., Levenez, F., Pons, N., Rasmussen, S., Sunagawa, S., Tap, J., Tims, S., Zoetendal, E. G., Brunak, S., Clément, K., Doré, J., Kleerebezem, M., Kristiansen, K., Renault, P., Sicheritz-Ponten, T., de Vos, W. M., Zucker, J.-D., Raes, J., Hansen, T., MetaHIT consortium, Bork, P., Wang, J., Ehrlich, S. D., and Pedersen, O. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature*, 500(7464):541–546.

[107] Lee, J., Yun, H., Feist, A. M., Palsson, B. Ø., and Lee, S. Y. (2008a). Genome-scale reconstruction and in silico analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. *Applied microbiology and biotechnology*, 80(5):849–862.

[108] Lee, J.-H. and O'Sullivan, D. J. (2010). Genomic insights into Bifidobacteria. *Microbiology and Molecular Biology Reviews*, 74(3):378–416.

[109] Lee, T. J., Paulsen, I., and Karp, P. (2008b). Annotation-based inference of transporter function. *Bioinformatics*, 24(13):i259–i267.

[110] Lewis, J. D., Chen, E. Z., Baldassano, R. N., Otley, A. R., Griffiths, A. M., Lee, D., Bittinger, K., Bailey, A., Friedman, E. S., Hoffmann, C., Albenberg, L., Sinha, R., Compher, C., Gilroy, E., Nessel, L., Grant, A., Chehoud, C., Li, H., Wu, G. D., and Bushman, F. D. (2015). Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell host & microbe*, 18(4):489–500.

[111] Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., Weitz, K. K., Eils, R., König, R., Smith, R. D., and Palsson, B. Ø. (2010). Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular systems biology*, 6(1):390.

[112] Lewis, N. E., Nagarajan, H., and Palsson, B. Ø. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature reviews. Microbiology*, 10(4):291.

[113] Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14):1754–1760.

[114] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

[115] Lien, K. A., McBurney, M. I., Beyde, B. I., Thomson, A., and Sauer, W. C. (1996). Ileal recovery of nutrients and mucin in humans fed total enteral formulas supplemented with soy fiber. *The American journal of clinical nutrition*, 63(4):584–595.

[116] Liu, J., Prindle, A., Humphries, J., Gabalda-Sagarra, M., Asally, M., yeon D Lee, D., Ly, S., Garcia-Ojalvo, J., and Süel, G. M. (2015). Metabolic codependence gives rise to collective oscillations within biofilms. *Nature*, 523(7562):550.

[117] Loferer-Krößbacher, M., Klima, J., and Psenner, R. (1998). Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Applied and Environmental Microbiology*, 64(2):688–694.

[118] Louca, S. and Doebeli, M. (2015). Calibration and analysis of genome-based models for microbial ecology. *Elife*, 4:e08208.

[119] Louis, P. and Flint, H. J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS microbiology letters*, 294(1):1–8.

[120] Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijs, I., Eeckhaut, V., Ballet, V., Claes, K., Van Immerseel, F., Verbeke, K., Ferrante, M., Verhaegen, J., Rutgeerts, P., and Vermeire, S. (2013). A decrease of the butyrate-producing species roseburia hominis and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut*, pages gutjnl–2013.

[121] Magnusdottir, S., Ravcheev, D., de Crecy-Lagard, V., and Thiele, I. (2015). Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Frontiers in genetics*, 6.

[122] Mahadevan, R. and Schilling, C. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic engineering*, 5(4):264–276.

[123] Manichaikul, A., Ghamsari, L., Hom, E. F., Lin, C., Murray, R. R., Chang, R. L., Balaji, S., Hao, T., Shen, Y., Chavali, A. K., Thiele, I., Yang, X., Fan, C., Mello, E., Hill, D. E., Vidal, M., Salehi-Ashtiani, K., and Papin, J. A. (2009). Metabolic network analysis integrated with transcript verification for sequenced genomes. *Nature methods*, 6(8):589–592.

[124] Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J., and Dore, J. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut*, 55(2):205–211.

[125] Marchandin, H., Teyssier, C., Campos, J., Jean-Pierre, H., Roger, F., Gay, B., Carlier, J.-P., and Jumas-Bilak, E. (2010). *Negativicoccus succinicivorans* gen. nov., sp. nov., isolated from human clinical samples, emended description of the family Veillonellaceae and description of Negativicutes classis nov., Selenomonadales ord. nov. and Acidaminococcaceae fam. nov. in the bacterial phylum Firmicutes. *International journal of systematic and evolutionary microbiology*, 60(6):1271–1279.

[126] Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N. N., and Kyrpides, N. C. (2011). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic acids research*, 40(D1):D115–D122.

[127] Maxwell, E. G., Belshaw, N. J., Waldron, K. W., and Morris, V. J. (2012). Pectin–an emerging new bioactive food polysaccharide. *Trends in Food Science & Technology*, 24(2):64–73.

[128] Mazumdar, V., Amar, S., and Segrè, D. (2013). Metabolic proximity in the order of colonization of a microbial community. *PloS one*, 8(10):e77617.

[129] McGuckin, M. A., Lindén, S. K., Sutton, P., and Florin, T. H. (2011). Mucin dynamics and enteric pathogens. *Nature reviews. Microbiology*, 9(4):265.

[130] MetaHIT Consortium (2011). MetaHIT: The european union project on metagenomics of the human intestinal tract. In *Metagenomics of the human body*, pages 307–316. Springer.

[131] Michaelis, L. and Menten, M. L. (2007). *Die Kinetik der Invertinwirkung*. Universitätsbibliothek Johann Christian Senckenberg.

[132] Middelboe, M. and Jørgensen, N. O. (2006). Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds. *Journal of the Marine Biological Association of the United Kingdom*, 86(3):605–612.

[133] Milne, C. B., Eddy, J. A., Raju, R., Ardekani, S., Kim, P.-J., Senger, R. S., Jin, Y.-S., Blaschek, H. P., and Price, N. D. (2011). Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. *BMC systems biology*, 5(1):130.

[134] Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2009). BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(suppl_1):D750–D753.

[135] Monk, J. M., Charusanti, P., Aziz, R. K., Lerman, J. A., Premyodhin, N., Orth, J. D., Feist, A. M., and Palsson, B. Ø. (2013). Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proceedings of the National Academy of Sciences*, 110(50):20338–20343.

[136] Monod, J. (1949). The growth of bacterial cultures. *Annual Reviews in Microbiology*, 3(1):371–394.

[137] Muralidharan, V., Rinker, K., Hirsh, I., Bouwer, E., and Kelly, R. (1997). Hydrogen transfer between methanogens and fermentative heterotrophs in hyperthermophilic cocultures. *Biotechnology and bioengineering*, 56(3):268–278.

[138] Nair, B., Mayberry, W., Dziak, R., Chen, P., Levine, M., and Hausmann, E. (1983). Biological effects of a purified lipopolysaccharide from *Bacteroides gingivalis*. *Journal of periodontal research*, 18(1):40–49.

[139] Natale, D. A., Shankavaram, U. T., Galperin, M. Y., Wolf, Y. I., Aravind, L., and Koonin, E. V. (2000). Towards understanding the first genome sequence of a crenarchaeon by genome annotation using clusters of orthologous groups of proteins (COGs). *Genome biology*, 1(5):research0009–1.

[140] Nookaew, I., Olivares-Hernández, R., Bhumiratana, S., and Nielsen, J. (2011). Genome-scale metabolic models of *Saccharomyces cerevisiae*. *Yeast Systems Biology: Methods and Protocols*, pages 445–463.

[141] O, H. C. V., Solheim, M., Snipen, L., Nes, I. F., and Brede, D. A. (2010). Comparative genomic analysis of pathogenic and probiotic *Enterococcus faecalis* isolates, and their transcriptional responses to growth in human urine. *PloS one*, 5(8):e12489.

[142] Oberhardt, M. A., Puchałka, J., Fryer, K. E., Dos Santos, V. M., and Papin, J. A. (2008). Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *Journal of bacteriology*, 190(8):2790–2803.

[143] O'brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., and Palsson, B. Ø. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular systems biology*, 9(1):693.

[144] Orth, J. D. and Palsson, B. Ø. (2012). Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions. *BMC systems biology*, 6(1):30.

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

[146] Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic acids research*, 33(17):5691–5702.

[147] Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., and Stevens, R. (2013). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic acids research*, 42(D1):D206–D214.

[148] Pamer, E. G. (2014). Fecal microbiota transplantation: effectiveness, complexities, and lingering concerns. *Mucosal immunology*, 7(2):210.

[149] Pande, S., Merker, H., Bohl, K., Reichelt, M., Schuster, S., De Figueiredo, L., Kaleta, C., and Kost, C. (2014). Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *The ISME journal*, 8(5):953.

[150] Parkins, M. D., Ceri, H., and Storey, D. G. (2001). *Pseudomonas aeruginosa* GacA, a factor in multihost virulence, is also essential for biofilm formation. *Molecular microbiology*, 40(5):1215–1226.

[151] Parvez, S., Malik, K. A., Ah Kang, S., and Kim, H.-Y. (2006). Probiotics and their fermented food products are beneficial for health. *Journal of applied microbiology*, 100(6):1171–1185.

[152] Peng, L., He, Z., Chen, W., Holzman, I. R., and Lin, J. (2007). Effects of butyrate on intestinal barrier function in a Caco-2 cell monolayer model of intestinal barrier. *Pediatric research*, 61(1):37–41.

[153] Plata, G., Henry, C. S., and Vitkup, D. (2015). Long-term phenotypic evolution of bacteria. *Nature*, 517(7534):369.

[154] Platzer, A. (2013). Visualization of SNPs with t-SNE. *PloS one*, 8(2):e56883.

[155] Prantera, C., Zannoni, F., Scribano, M. L., Berto, E., Andreoli, A., Kohn, A., and Luzi, C. (1996). An antibiotic regimen for the treatment of active Crohn's disease: a randomized, controlled clinical trial of metronidazole plus ciprofloxacin. *American Journal of Gastroenterology*, 91(2).

[156] Prats, R. and De Pedro, M. (1989). Normal growth and division of *Escherichia coli* with a reduced amount of murein. *Journal of bacteriology*, 171(7):3740–3745.

[157] Prentice, M. B. (2004). Bacterial comparative genomics. *Genome biology*, 5(8):338.

[158] Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Paslier, D. L., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., MetaHIT Consortium, Bork, P., Ehrlich, S. D., and Wang, J. (2010). A human gut microbial gene catalog established by metagenomic sequencing. *Nature*, 464(7285):59.

[159] Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., Peng, Y., Zhang, D., Jie, Z., Wu, W., Qin, Y., Xue, W., Li, J., Han, L., Lu, D., Wu, P., Dai, Y., Sun, X., Li, Z., Tang, A., Zhong, S., Li, X., Chen, W., Xu, R., Wang, M., Feng, Q., Gong, M., Yu, J., Zhang, Y., Zhang, M., Hansen, T., Sanchez, G., Raes, J., Falony, G., Okuda, S., Almeida, M., LeChatelier, E., Renault, P., Pons, N., Batto, J.-M., Zhang, Z., Chen, H., Yang, R., Zheng, W., Li, S., Yang, H., Wang, J., Ehrlich, S. D., Nielsen,

R., Pedersen, O., Kristiansen, K., and Wang, J. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, 490(7418):55–60.

[160] Reed, J. L., Patel, T. R., Chen, K. H., Joyce, A. R., Applebee, M. K., Herring, C. D., Bui, O. T., Knight, E. M., Fong, S. S., and Palsson, B. Ø. (2006). Systems approach to refining genome annotation. *Proceedings of the National Academy of Sciences*, 103(46):17480–17484.

[161] Robert, M., Mercade, M., Bosch, M., Parra, J., Espuny, M., Manresa, M., and Guinea, J. (1989). Effect of the carbon source on biosurfactant production by *Pseudomonas aeruginosa* 44T1. *Biotechnology Letters*, 11(12):871–874.

[162] Roediger, W. (1982). Utilization of nutrients by isolated epithelial cells of the rat colon. *Gastroenterology*, 83(2):424–429.

[163] Rolfsson, O., Paglia, G., Magnusdottir, M., Palsson, B. Ø., and Thiele, I. (2013). Inferring the metabolism of human orphan metabolites from their metabolic network context affirms human gluconokinase activity. *Biochemical Journal*, 449(2):427–435.

[164] Sabatino, A., Morera, R., Ciccocioppo, R., Cazzola, P., Gotti, S., Tinozzi, F. P., Tinozzi, S., and Corazza, G. R. (2005). Oral butyrate for mildly to moderately active Crohn's disease. *Alimentary pharmacology & therapeutics*, 22(9):789–794.

[165] Salem, H., Bauer, E., Strauss, A. S., Vogel, H., Marz, M., and Kaltenpoth, M. (2014). Vitamin supplementation by gut symbionts ensures metabolic homeostasis in an insect host. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1796):20141838.

[166] Sangwan, N., Xia, F., and Gilbert, J. A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome*, 4(1):8.

[167] Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., Kang, J., Hyduke, D. R., and Palsson, B. Ø. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature protocols*, 6(9):1290.

[168] Schluter, J. and Foster, K. R. (2012). The evolution of mutualism in gut microbiota via host epithelial selection. *PLoS biology*, 10(11):e1001424.

[169] Schobert, M. and Jahn, D. (2010). Anaerobic physiology of *Pseudomonas aeruginosa* in the cystic fibrosis lung. *International Journal of Medical Microbiology*, 300(8):549–556.

[170] Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., and Schomburg, D. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic acids research*, 32(suppl_1):D431–D433.

[171] Segata, N., Börnigen, D., Morgan, X. C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature communications*, 4:2304.

[172] Sengupta, S., Muir, J. G., and Gibson, P. R. (2006). Does butyrate protect from colorectal cancer? *Journal of gastroenterology and hepatology*, 21(1):209–218.

[173] Shah, P., Fritz, J. V., Glaab, E., Desai, M. S., Greenhalgh, K., Frachet, A., Niegowska, M., Estes, M., Jäger, C., Seguin-Devaux, C., Zenhausern, F., and Wilmes, P. (2016). A microfluidics-based in vitro model of the gastrointestinal human–microbe interface. *Nature communications*, 7.

[174] Shashkova, T., Popenko, A., Tyakht, A., Peskov, K., Kosinsky, Y., Bogolubsky, L., Raigorodskii, A., Ischenko, D., Alexeev, D., and Govorun, V. (2016). Agent based modeling of human gut microbiome interactions and perturbations. *PloS one*, 11(2):e0148386.

[175] Shoaie, S., Ghaffari, P., Kovatcheva-Datchary, P., Mardinoglu, A., Sen, P., Pujos-Guillot, E., de Wouters, T., Juste, C., Rizkalla, S., Chilloux, J., L, H., K, N. J., MICRO-Obes Consortium, J, D., E, D. M., K, C., F, B., and J, N. (2015). Quantifying diet-induced metabolic changes of the human gut microbiome. *Cell metabolism*, 22(2):320–331.

[176] Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*, 480(7376):241.

[177] Smirnov, A., Sklan, D., and Uni, Z. (2004). Mucin dynamics in the chick small intestine are altered by starvation. *The Journal of nutrition*, 134(4):736–742.

[178] Smyth, P. F. and Clarke, P. H. (1975). Catabolite repression of *Pseudomonas aeruginosa* amidase: the effect of carbon source on amidase synthesis. *Microbiology*, 90(1):81–90.

[179] Soetaert, K. and Meysman, F. (2010). ReacTran: Reactive transport modelling in 1D, 2D and 3D. *R package version*, 1.

[180] Soetaert, K., Petzoldt, T., and Setzer, R. W. (2010). Solving differential equations in R: package deSolve. *Journal of Statistical Software*, 33.

[181] Sokal, R. R. and Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon*, 11(2):33–40.

[182] Stearns, J. C., Lynch, M. D., Senadheera, D. B., Tenenbaum, H. C., Goldberg, M. B., Cvitkovitch, D. G., Croitoru, K., Moreno-Hagelsieb, G., and Neufeld, J. D. (2011). Bacterial biogeography of the human digestive tract. *Scientific reports*, 1:170.

[183] Stefania Magnusdottir, Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., Wilmes, P., Fleming, R. M. T., and Thiele, I. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature biotechnology*, 35(1):81–89.

[184] Stewart, P. S. (2003). Diffusion in biofilms. *Journal of bacteriology*, 185(5):1485–1491.

[185] Stewart, P. S. and Costerton, J. W. (2001). Antibiotic resistance of bacteria in biofilms. *The lancet*, 358(9276):135–138.

[186] Stewart, P. S. and Franklin, M. J. (2008). Physiological heterogeneity in biofilms. *Nature reviews. Microbiology*, 6(3):199.

[187] Sung, J., Kim, S., Cabatbat, J. J. T., Jang, S., Jin, Y.-S., Jung, G. Y., Chia, N., and Kim, P.-J. (2017). Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *arXiv preprint arXiv:1706.01787*.

[188] Suthers, P. F., Dasika, M. S., Kumar, V. S., Denisov, G., Glass, J. I., and Maranas, C. D. (2009). A genome-scale metabolic reconstruction of *Mycoplasma genitalium*, iPS189. *PLoS Computational Biology*, 5(2):e1000285.

[189] Swidsinski, A., Weber, J., Loening-Baucke, V., Hale, L. P., and Lochs, H. (2005). Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *Journal of clinical microbiology*, 43(7):3380–3389.

[190] Tegtmeier, D., Thompson, C. L., Schauer, C., and Brune, A. (2016). Oxygen affects gut bacterial colonization and metabolic activities in a gnotobiotic cockroach model. *Applied and Environmental Microbiology*, 82(4):1080–1089.

[191] Thiele, I., Heinken, A., and Fleming, R. M. T. (2013a). A systems biology approach to studying the role of microbes in human health. *Current opinion in biotechnology*, 24(1):4–12.

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

[193] Thiele, I., Swainston, N., Fleming, R. M. T., Hoppe, A., Sahoo, S., Aurich, M. K., Haraldsdottir, H., Mo, M. L., Rolfsson, O., Stobbe, M. D., Thorleifsson, S. G., Agren, R., Bölling, C., Bordel, S., Chavali, A. K., Dobson, P., Dunn, W. B., Endler, L., Hala, D., Hucka, M., Hull, D., Jameson, D., Jamshidi, N., Jonsson, J. J., Juty, N., Keating, S., Nookaew, I., Novere, N. L., Malys, N., Mazein, A., Papin, J. A., Price, N. D., Sr, E. S., Sigurdsson, M. I., Simeonidis, E., Sonnenschein, N., Smallbone, K., Sorokin, A., van Beek, J. H. G. M., Weichart, D., Goryanin, I., Nielsen, J., Westerhoff, H. V., Kell, D. B., Mendes, P., and Palsson, B. Ø. (2013b). A community-driven global reconstruction of human metabolism. *Nature biotechnology*, 31(5):419–425.

[194] Thiele, I., Vlassis, N., and Fleming, R. M. T. (2014). fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics*, 30(17):2529–2531.

[195] Thomas, G. H., Zucker, J., Macdonald, S. J., Sorokin, A., Goryanin, I., and Douglas, A. E. (2009). A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *BMC systems biology*, 3(1):24.

[196] Tilg, H. and Kaser, A. (2011). Gut microbiome, obesity, and metabolic dysfunction. *The Journal of clinical investigation*, 121(6):2126.

[197] Tremaroli, V. and Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature*, 489(7415):242.

[198] Trosvik, P. and Muinck, E. J. (2015). Ecology of bacteria in the human gastrointestinal tract—identification of keystone and foundation taxa. *Microbiome*, 3(1):44.

[199] Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C., Knight, R., and Gordon, J. I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature*, 449(7164):804.

[200] Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., and Gordon, J. I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Science translational medicine*, 1(6):6ra14–6ra14.

[201] Turroni, F., Ribbera, A., Foroni, E., van Sinderen, D., and Ventura, M. (2008). Human gut microbiota and Bifidobacteria: from composition to functionality. *Antonie Van Leeuwenhoek*, 94(1):35–50.

[202] van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605.

[203] van Dullemen, H. M., van Deventer, S. J., Hommes, D. W., Bijl, H. A., Jansen, J., Tytgat, G. N., and Woody, J. (1995). Treatment of Crohn's disease with anti-tumor necrosis factor chimeric monoclonal antibody (cA2). *Gastroenterology*, 109(1):129–135.

[204] van Hoek, M. J. A. and Merks, R. M. H. (2017). Emergence of microbial diversity due to cross-feeding interactions in a spatial model of gut microbial metabolism. *BMC systems biology*, 11(1):56.

[205] van Passel, M. W., Kant, R., Zoetendal, E. G., Plugge, C. M., Derrien, M., Malfatti, S. A., Chain, P. S., Woyke, T., Palva, A., de Vos, W. M., and Smidt, H. (2011). The genome of *Akkermansia muciniphila*, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. *PloS one*, 6(3):e16876.

[206] Vander Wauven, C., Pierard, A., Kley-Raymann, M., and Haas, D. (1984). *Pseudomonas aeruginosa* mutants affected in anaerobic growth on arginine: evidence for a four-gene cluster encoding the arginine deiminase pathway. *Journal of bacteriology*, 160(3):928–934.

[207] Vos, P., Garrity, G., Jones, D., Krieg, N. R., Ludwig, W., Rainey, F. A., Schleifer, K.-H., and Whitman, W. (2011). *Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*, volume 3. Springer Science & Business Media.

[208] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83.

[209] Wilmes, P., Bowen, B. P., Thomas, B. C., Mueller, R. S., Denef, V. J., VerBerkmoes, N. C., Hettich, R. L., Northen, T. R., and Banfield, J. F. (2010). Metabolome-proteome differentiation coupled to microbial divergence. *MBio*, 1(5):e00246–10.

[210] Wilschanski, M., Sherman, P., Pencharz, P., Davis, L., Corey, M., and Griffiths, A. (1996). Supplementary enteral nutrition maintains remission in paediatric Crohn's disease. *Gut*, 38(4):543–548.

[211] Wolkenhauer, O. (2014). Why model? *Frontiers in physiology*, 5.

[212] Wong, J. M., De Souza, R., Kendall, C. W., Emam, A., and Jenkins, D. J. (2006). Colonic health: fermentation and short chain fatty acids. *Journal of clinical gastroenterology*, 40(3):235–243.

[213] Yacoubi, B. E. and de Crecy-Lagar, V. (2014). Integrative data-mining tools to link gene and function. *Gene Function Analysis*, pages 43–66.

[214] Yizhak, K., Tuller, T., Papp, B., and Ruppin, E. (2011). Metabolic modeling of endosymbiont genome reduction on a temporal scale. *Molecular systems biology*, 7(1):479.

[215] Yutin, N. and Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced Clostridia. *Environmental microbiology*, 15(10):2631–2641.

[216] Zaneveld, J. R., Lozupone, C., Gordon, J. I., and Knight, R. (2010). Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. *Nucleic acids research*, 38(12):3869–3879.

[217] Zengler, K. and Palsson, B. Ø. (2012). A road map for the development of community systems (CoSy) biology. *Nature reviews. Microbiology*, 10(5):366.

[218] Zhang, H., DiBaise, J. K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., Parameswaran, P., Crowell, M. D., Wing, R., Rittmann, B. E., and Krajmalnik-Brown, R. (2009). Human gut microbiota in obesity and after gastric bypass. *Proceedings of the National Academy of Sciences*, 106(7):2365–2370.

[219] Zhang, H., Gao, S., Lercher, M. J., Hu, S., and Chen, W.-H. (2012). EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic acids research*, 40(W1):W569–W572.

[220] Zhu, C., Delmont, T. O., Vogel, T. M., and Bromberg, Y. (2015). Functional basis of microorganism classification. *PLoS computational biology*, 11(8):e1004472.

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

[222] Zoetendal, E., Rajilić-Stojanović, M., and De Vos, W. (2008). High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut*, 57(11):1605–1615.

[223] Zoetendal, E. G., Raes, J., Van Den Bogert, B., Arumugam, M., Booijink, C. C., Troost, F. J., Bork, P., Wels, M., De Vos, W. M., and Kleerebezem, M. (2012). The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *The ISME journal*, 6(7):1415.

[224] Zomorrodi, A. R., Islam, M. M., and Maranas, C. D. (2014). d-OptCom: dynamic multi-level and multi-objective metabolic modeling of microbial communities. *ACS synthetic biology*, 3(4):247–257.

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

[226] Zomorrodi, A. R. and Segrè, D. (2016). Synthetic ecology of microbes: mathematical models and applications. *Journal of molecular biology*, 428(5):837–861.

Appendix A

Supplementary material for Chapter 2

A.1 Supplementary tables

The following tables are too large to be displayed in text and are available via the publisher's website.

Table A.1: Table of the gap-filled reactions used to ensure anaerobic growth. Direct download: https://static-content.springer.com/esm/art%3A10.1186%2Fs40168-015-0121-6/MediaObjects/40168_2015_121_MOESM1_ESM.xlsx

Columns	Description
SEED.ID	Model ID automatically assigned by the SEED database.
Reaction Added	IDs of reactions that were added to the respective model.
Reaction Descriptions	IDs of reactions that were added to the respective model.
Formula	Reaction formula of the gapfilled reactions with the respective metabolites.

Table A.2: List of genome and model statistics of the microbe selection. Direct download: https://static-content.springer.com/esm/art%3A10.1186%2Fs40168-015-0121-6/MediaObjects/40168_2015_121_MOESM3_ESM.xlsx

Columns	Description
SEED.ID	Model ID automatically assigned by the SEED database.
IMG.Genome.ID	Model ID automatically assigned by the IMG database.
NCBI.Taxon.ID	Model ID automatically assigned by the NCBI taxonomy database.
taxon oid	Taxon ID assigned by SEED.
Status	Genome sequencing status.
Study.Name	Name of the study under IMG database.
Organism	Organism name based on the IMG database.
Domain	Taxonomic domain of organisms.
Phylum	Taxonomic phylum of organisms.
Class	Taxonomic class of organisms.
Order	Taxonomic order of organisms.
Family	Taxonomic family of organisms.
Genus	Taxonomic genus of organisms.
Species	Taxonomic species of organisms.
Strain	Taxonomic strain of organisms.
Genome.Size	Genome size of organisms.

Table A.3: List of all reactions sorted according to their contribution to the point separation. Direct download: https://static-content.springer.com/esm/art%3A10.1186%2Fs40168-015-0121-6/MediaObjects/40168_2015_121_MOESM4_ESM.xlsx

Columns	Description
SEED ID	Reaction ID automatically assigned by the SEED database.
BiGG ID	IDs of reactions based on BiGG nomenclature.
Reaction formula	Reaction formula of the reactions with the respective metabolites.
Correlation	Sorted absolute value of eigenvalues for each reaction.

Table A.4: List of genera members belonging to the different clusters presented. Direct download: https://static-content.springer.com/esm/art%3A10.1186%2Fs40168-015-0121-6/MediaObjects/40168_2015_121_MOESM9_ESM.xlsx

Columns	Description
Organism	Name of the strains.
Cluster	Cluster name to which each organism belongs.

Table A.5: Table with reaction differences within the clusters found for Bifidobacterium. Direct download: https://static-content.springer.com/esm/art%3A10.1186%2Fs40168-015-0121-6/MediaObjects/40168_2015_121_MOESM11_ESM.xlsx

Columns	Description
SEED ID	Reaction ID automatically assigned by the SEED database.
BiGG ID	IDs of reactions based on BiGG nomenclature.
Reaction formula	Reaction formula of the reactions with the respective metabolites.
Description	IDs of reactions that were added to the respective model.
Reaction formula	Reaction formula of the gapfilled reactions with the respective metabolites.
Cluster 1 Number of Organisms	Number of organism in cluster 1 that have the corresponding reaction.
Cluster 2 Number of Organisms	Number of organism in cluster 2 that have the corresponding reaction.

Table A.6: Table with reaction differences within the clusters found for Bacteroides. Direct download: https://static-content.springer.com/esm/art%3A10.1186%2Fs40168-015-0121-6/MediaObjects/40168_2015_121_MOESM12_ESM.xlsx

Columns	Description
SEED ID	Reaction ID automatically assigned by the SEED database.
BiGG ID	IDs of reactions based on BiGG nomenclature.
Reaction formula	Reaction formula of the reactions with the respective metabolites.
Description	IDs of reactions that were added to the respective model.
Reaction formula	Reaction formula of the gapfilled reactions with the respective metabolites.
Cluster 1 Number of Organisms	Number of organism in cluster 1 that have the corresponding reaction.
Cluster 2 Number of Organisms	Number of organism in cluster 2 that have the corresponding reaction.

Table A.7: The fitted parameters of the exponential models

Figure	alpha	beta
Figure 4 A	-0.027	1.967
Figure 4 B	0.002	0.145
Figure 4 C	-0.052	1.824
Figure 4 D	-0.144	1.95

A.2 Supplementary figures

A

Model	Reactions in curated model	Reactions in draft model	Overlapping reactions	Reactions only in curated model	Reactions only in draft model
Bacteroides thetaiotaomicron	1528	1014	780	748	234
Faecalibacterium prausnitzii	1030	891	610	420	281
Lactobacillus plantarum WCFS1	777	1052	341	436	711
Streptococcus thermophilus LMG18311	556	833	277	279	556
Helicobacter pylori 26695	555	877	263	292	585
Klebsiella pneumoneae MGH 78578	2262	1621	765	1497	856
Salmonella enterica subsp. typhimurium LT2	2623	1575	768	1855	789
Escherichia coli MG1655	2426	1619	777	1649	842
E. coli O157:H7 strain Sakai	2372	1583	752	1620	831

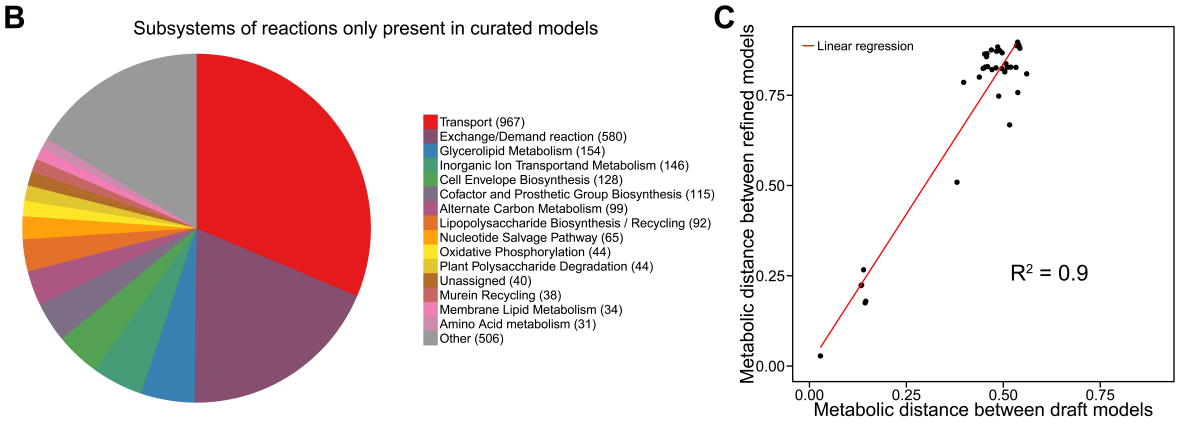

Figure A.1: Comparison between a set of our draft reconstructions and a set of published manually curated reconstructions.

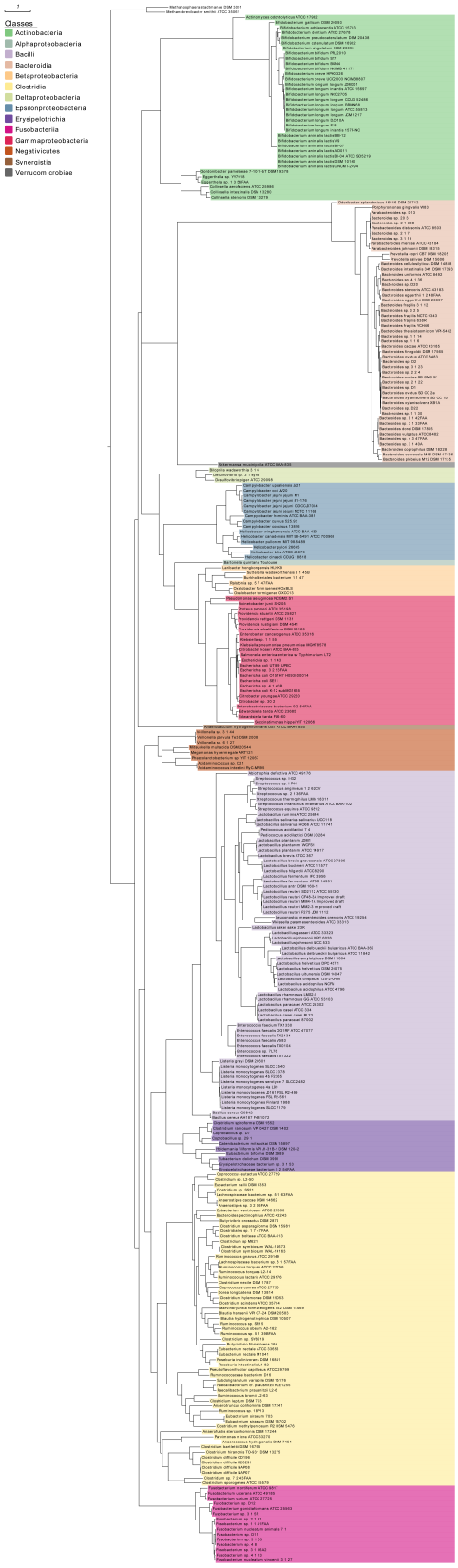

Figure A.2: Phylogenetic maximum likelihood tree (rooted with two methanogenic archaea) calculated from the sequence similarity of 400 selected essential genes.

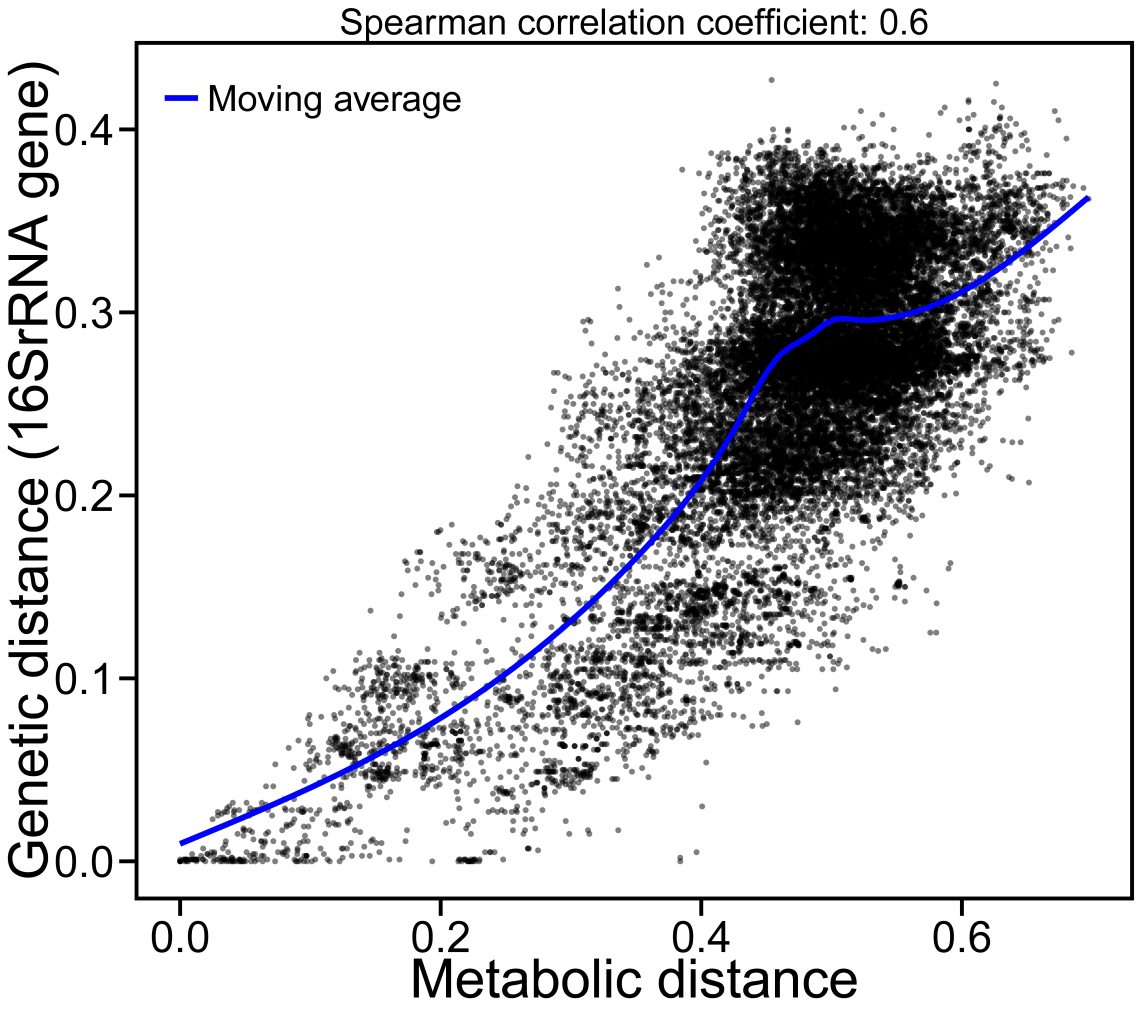

Figure A.3: The exponential relationship between the phylogeny and reaction content using the 16S rRNA sequence similarity as a measure for genetic distance.

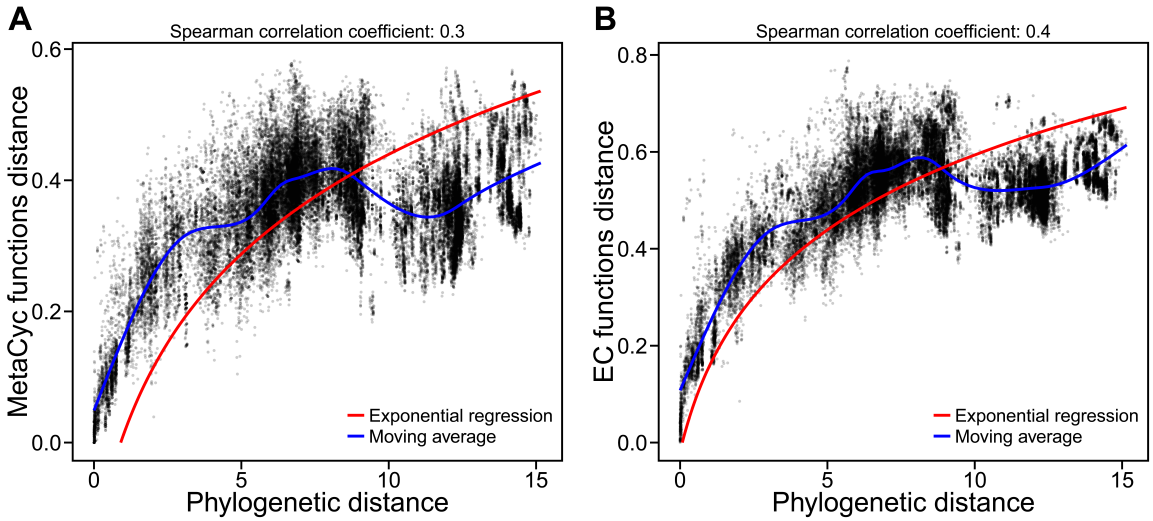

Figure A.4: The correlation between MetaCyc and EC functionalities with the phylogenetic distance.

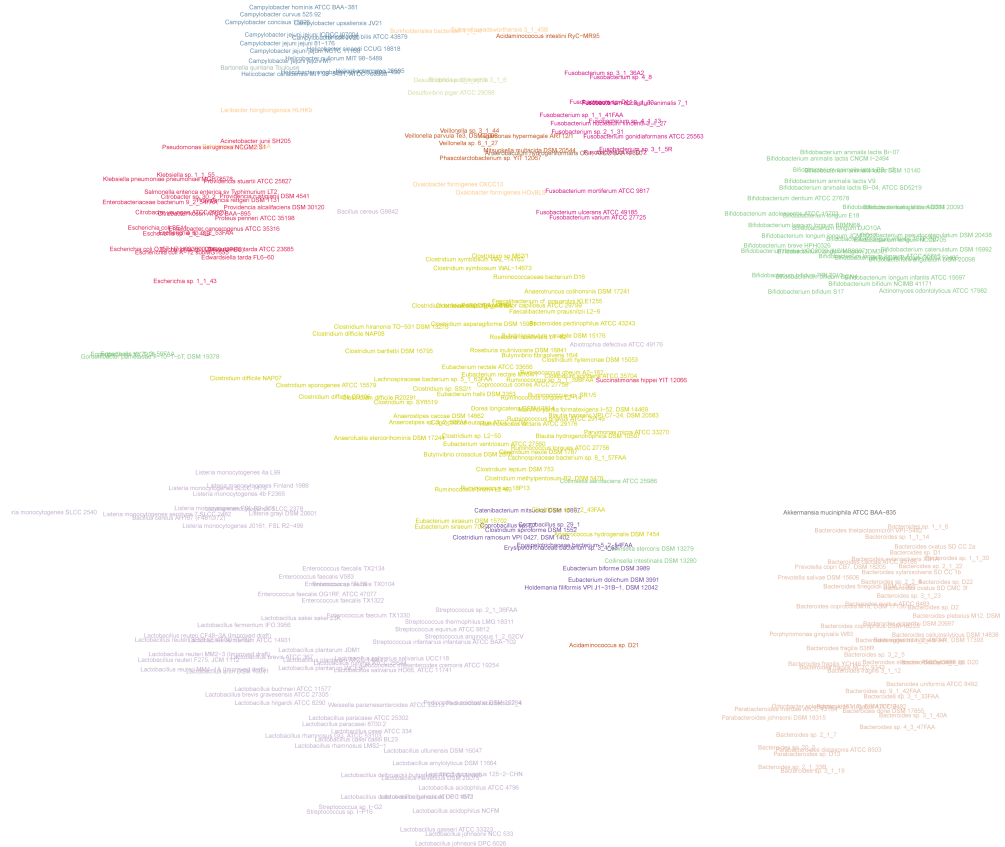

Figure A.5: The same t-SNE-based, two-dimensional coordinates as in Figure 5 with additional point labels for the different organisms.

Appendix B

Supplementary material for Chapter 3

B.1 Supplementary tables

Table B.1: Table of all exchange reactions with their respective concentrations that were added to the environment.

Substance	Reaction.ID	Concentration	diffusion.constant
Orthophosphate	EX_EC0009	0.10	0.00
Cobalt	EX_EC0144	0.10	0.00
H2O	EX_EC0001	0.10	0.00
Magnesium	EX_EC0248	0.10	0.00
Nitrogen	EX_EC0518	0.10	0.00
Zinc	EX_EC0034	0.10	0.00
Sodium	EX_EC0954	0.10	0.00
Cadmium	EX_EC0994	0.10	0.00
Copper	EX_EC0056	0.10	0.00
NH4+	EX_EC0957	0.10	0.00
CO2	EX_EC0011	0.10	0.00
D-Glucose	EX_EC0027	0.05	0.00
Sulfate	EX_EC0048	0.10	0.00
Manganese	EX_EC0030	0.10	0.00
Iron	EX_EC0021	0.10	0.00
H+	EX_EC0065	0.10	0.00
Potassium	EX_EC0197	0.10	0.00
Oxygen	EX_EC0007	0.10	0.00

Table B.2: Table of all exchange reactions of the defined essential metabolites, mucus glycans, and remaining metabolites with their respective concentrations. Direct download: https://doi.org/10.1371/journal.pcbi.1005544.s009

Columns	Description
Exchange	Identifier of exchange reaction in the model.
diffconstant	Diffusion constant for each metabolite of the corresponding exchange reaction.
concentration im mM	Concentration for each metabolite of the corresponding exchange reaction.
category	Category for each metabolite of the corresponding exchange reaction.

B.2 Supplementary figures

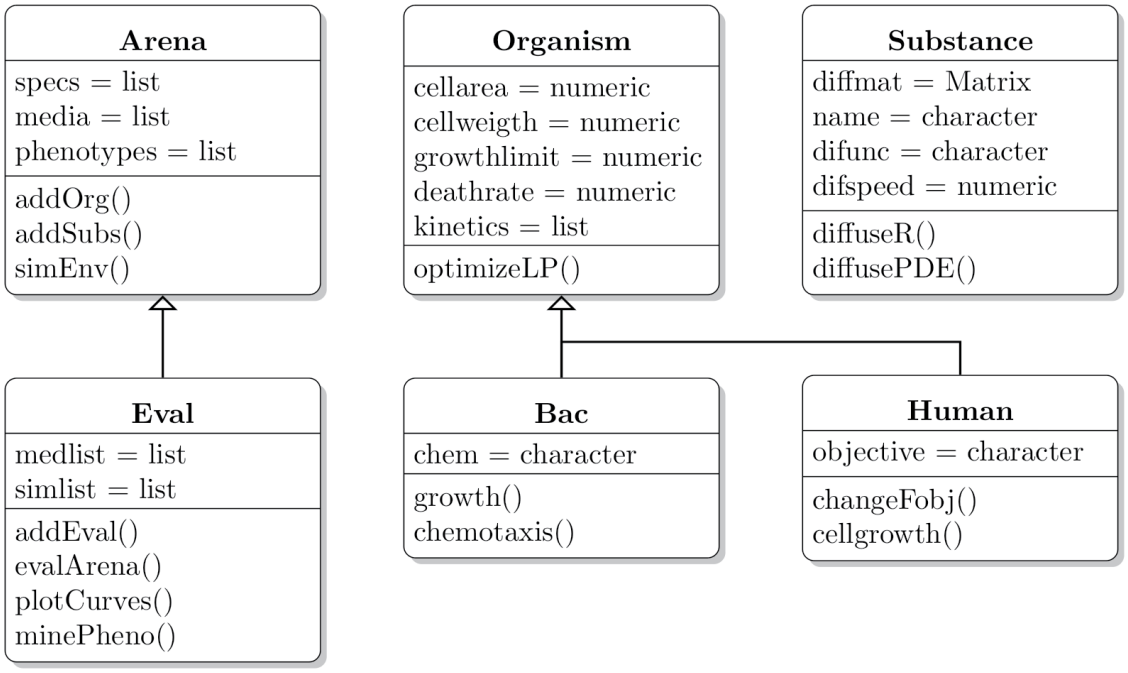

Figure B.1: Simplified class diagram displaying the inheritance hierarchy.

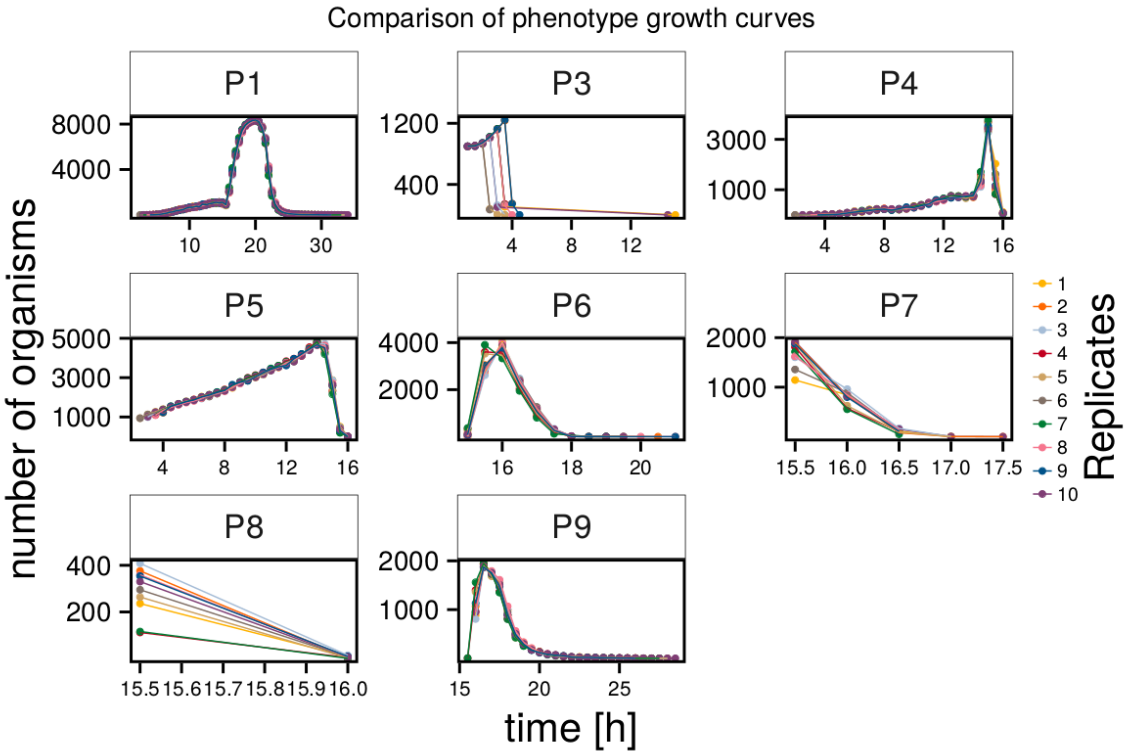

Figure B.2: For each phenotype (P2,P3,. . . ,P9) of the P. aeruginosa biofilm simulation the time curves for all replicates are shown. While the overall dynamics were stable, the occurrences of P3, P7 and P8 showed some minor variance.

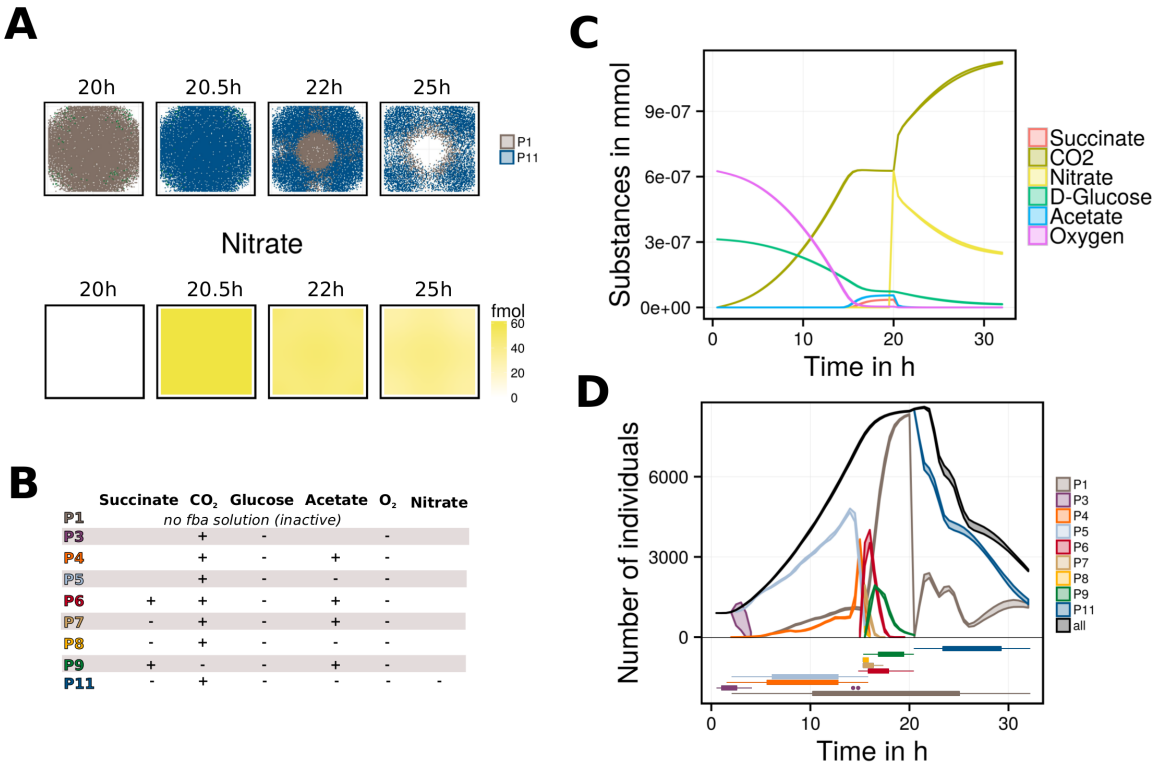

Figure B.3: Alternative scenario of P. aeruginosa biofilm simulation with 0.1 mM nitrate added after 20 hours simulation time. A Spatial distribution of phenotypes and nitrate. The presence of nitrate after 20 hours was accomplished by a new nitrate consuming phenotype P11. B Comparison of phenotypes. C Time curve of core metabolites. The addition of nitrate after 20 hours lead to further glucose usage and CO2 production. The former produced acetate and succinate were used again. D Phenotypes growth curve. After the addition of nitrate, the metabolic inactive phenotype P1 vanished and the new nitrate consuming phenotype P11 emerged.

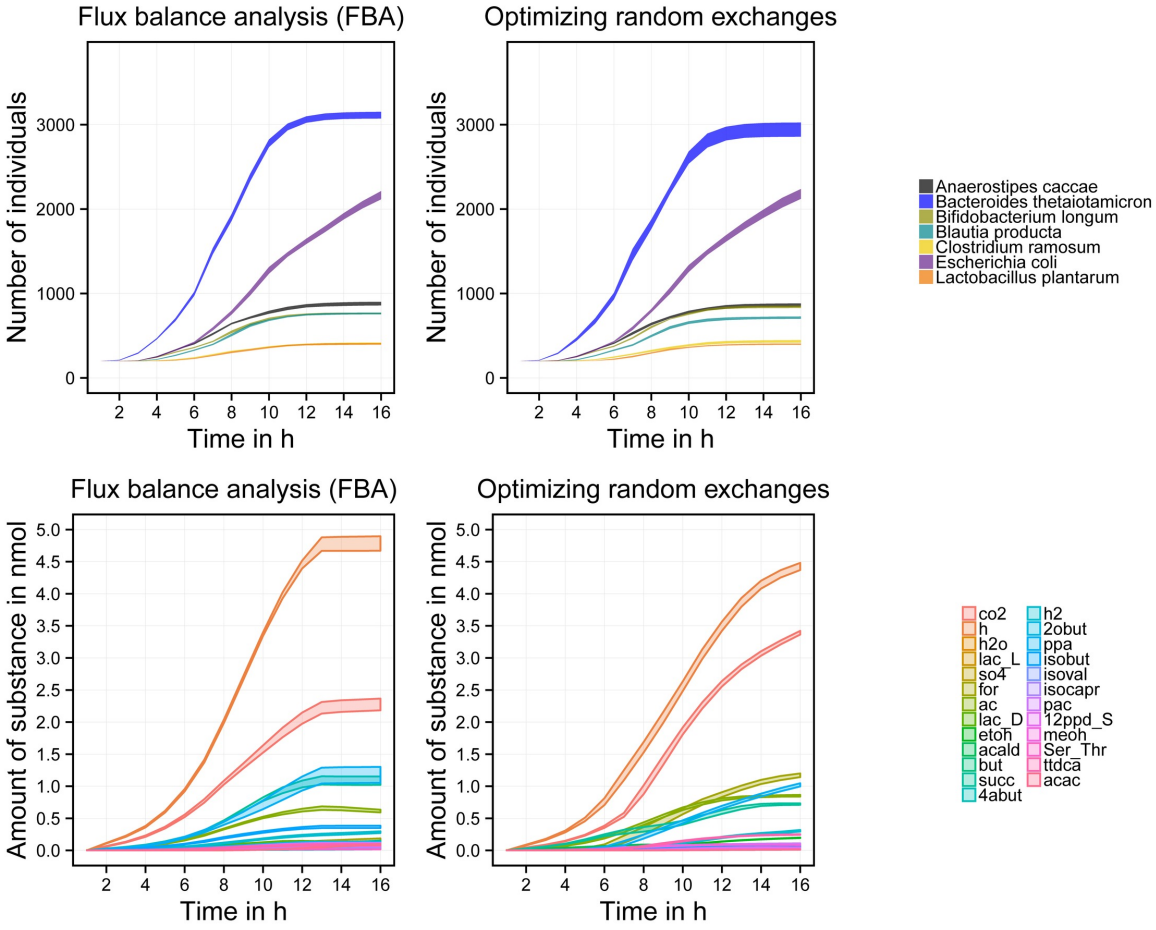

Figure B.4: The first row represents the species growth and the second row the concentration change of the 25 most variable metabolites. The first columns shows a default flux balance analysis and the second column the optimization of a random exchange reaction as a secondary objective. The curve range shows a standard deviation of 10 replicate simulations each simulating 16 hours.

B.3 Supplementary notes

The following notes are too large to be displayed in text and are available via the publisher's website.

B.3.1 Tutorial for BacArena

This tutorial includes a basic hands-on description of all main classes and functions of BacArena. Direct download: https://doi.org/10.1371/journal.pcbi.1005544.s001

B.3.2 Reference manual of BacArena.

All methods and parameters are explained with words and example codes in the documentation. Direct download: https://doi.org/10.1371/journal.pcbi.1005544.s002

B.3.3 P. aeruginosa single-species biofilm.

Documentation of changes in metabolic model of P. aeruginosa and additional figures from replicates. Direct download: https://doi.org/10.1371/journal.pcbi.1005544.s003

B.4 Supplementary files

The following files are too large to be displayed in text and are available via the publisher's website.

B.4.1 R Data file of modified P. aeruginosa model.

Metabolic model of P. aeruginosa used in simulation. Direct download: `https://doi.org/10.1371/journal.pcbi.1005544.s004`

B.4.2 R script to reproduce P. aeruginosa simulation.

This R script reproduces the biofilm simulation of P. aeruginosa simulation. Direct download: `https://doi.org/10.1371/journal.pcbi.1005544.s005`

B.4.3 R Data file with all 7 species used for the gut simulation.

Metabolic models of A. caccae, B. thetaiotaomicron, B. producta, E. coli, C. ramosum, L. plantarum, B. longum, and A. muciniphila used for the simulation of a simplified human gut model. Direct download: `https://doi.org/10.1371/journal.pcbi.1005544.s006`

B.4.4 R script to reproduce gut simulation.

The R script can be used to reproduce the gut-community simulation. Direct download: `https://doi.org/10.1371/journal.pcbi.1005544.s007`

Appendix C

Supplementary material for Chapter 4

C.1 Supplementary tables

Table C.1: List of 20 reactions most contributing to the point separation of the reaction differences between patients.

Reaction ID	MDS1	MDS2	Subsystem
MPPP9MT	0.08	0.01	Vitamin B12 metabolism
r0556c	0.08	0.02	Citric acid cycle
CELLUL_DEGe	0.08	0.02	Plant polysaccharide degradation
EX_cellul(e)	0.08	0.02	Exchange/demand reaction
EX_pppi(e)	0.08	0.02	Exchange/demand reaction
SO3rDmq	0.08	0.02	Sulfur metabolism
HXANt2r	0.08	0.02	Transport, extracellular
SO3rDdmq	0.08	0.02	Sulfur metabolism
FDX_NAD_NADP_OXi	0.08	0.01	Sulfur metabolism
SO3R	0.08	0.01	Sulfur metabolism
ADXFTDA	0.08	0.02	Ubiquinone and other terpenoid-quinone biosynthesis
CHDHR	0.08	0.02	Ubiquinone and other terpenoid-quinone biosynthesis
FTHD	0.08	0.02	Ubiquinone and other terpenoid-quinone biosynthesis
SACALDACT	0.08	0.02	Sulfur metabolism
TAURPYRAT	0.08	0.02	Sulfur metabolism
H2O2syn	0.08	0.01	Tyrosine metabolism
NOX1	0.08	0.01	Energy metabolism
2MBUTt2r	0.08	0.02	Transport, extracellular
5APTNt2r	0.08	0.02	Transport, extracellular
ALAD_L	0.08	0.02	Alanine and aspartate metabolism

Table C.2: Patient metagenomic data accession numbers with the respective categories.

ReadAccession	ExperimentAccession	Disease	ID
SRR2145575	SRX1133436	Crohn's disease	CD1
SRR2145587	SRX1133448	Crohn's disease	CD2
SRR2145609	SRX1133468	Crohn's disease	CD3
SRR2145607	SRX1133472	Crohn's disease	CD4
SRR2145623	SRX1133484	Crohn's disease	CD5
SRR2145635	SRX1133496	Crohn's disease	CD6
SRR2145508	SRX1133511	Crohn's disease	CD7
SRR2145524	SRX1133527	Crohn's disease	CD8
SRR2145528	SRX1133531	Crohn's disease	CD9
SRR2145544	SRX1133551	Crohn's disease	CD10
SRR2145392	SRX1133583	Crohn's disease	CD11
SRR2145400	SRX1133591	Crohn's disease	CD12
SRR2145412	SRX1133603	Crohn's disease	CD13
SRR2145416	SRX1133607	Crohn's disease	CD14
SRR2145420	SRX1133611	Crohn's disease	CD15
SRR2145428	SRX1133619	Crohn's disease	CD16
SRR2145436	SRX1133627	Crohn's disease	CD17
SRR2145440	SRX1133631	Crohn's disease	CD18
SRR2145448	SRX1133639	Crohn's disease	CD19
SRR2145548	SRX1133651	Crohn's disease	CD20
SRR2145552	SRX1133655	Crohn's disease	CD21
SRR2145315	SRX1133666	Crohn's disease	CD22
SRR2145323	SRX1133674	Crohn's disease	CD23
SRR2145331	SRX1133685	Crohn's disease	CD24
SRR2145335	SRX1133689	Crohn's disease	CD25
SRR2145467	SRX1133697	Crohn's disease	CD26
SRR2145471	SRX1133701	Crohn's disease	CD27
SRR2145483	SRX1133713	Crohn's disease	CD28
SRR2145359	SRX1133410	Control	HC1
SRR2145360	SRX1133411	Control	HC2
SRR2145361	SRX1133412	Control	HC3
SRR2145362	SRX1133413	Control	HC4
SRR2145363	SRX1133414	Control	HC5
SRR2145364	SRX1133415	Control	HC6
SRR2145365	SRX1133416	Control	HC7
SRR2145366	SRX1133417	Control	HC8
SRR2145367	SRX1133418	Control	HC9
SRR2145368	SRX1133419	Control	HC10
SRR2145369	SRX1133420	Control	HC11
SRR2145370	SRX1133421	Control	HC12
SRR2145371	SRX1133422	Control	HC13
SRR2145372	SRX1133423	Control	HC14
SRR2145373	SRX1133424	Control	HC15
SRR2145374	SRX1133425	Control	HC16
SRR2145375	SRX1133426	Control	HC17
SRR2145376	SRX1133427	Control	HC18
SRR2145377	SRX1133428	Control	HC19
SRR2145378	SRX1133429	Control	HC20
SRR2145379	SRX1133430	Control	HC21
SRR2145380	SRX1133431	Control	HC22
SRR2145381	SRX1133432	Control	HC23
SRR2145382	SRX1133433	Control	HC24
SRR2145383	SRX1133434	Control	HC25
SRR2145384	SRX1133435	Control	HC26

C.2 Supplementary figures

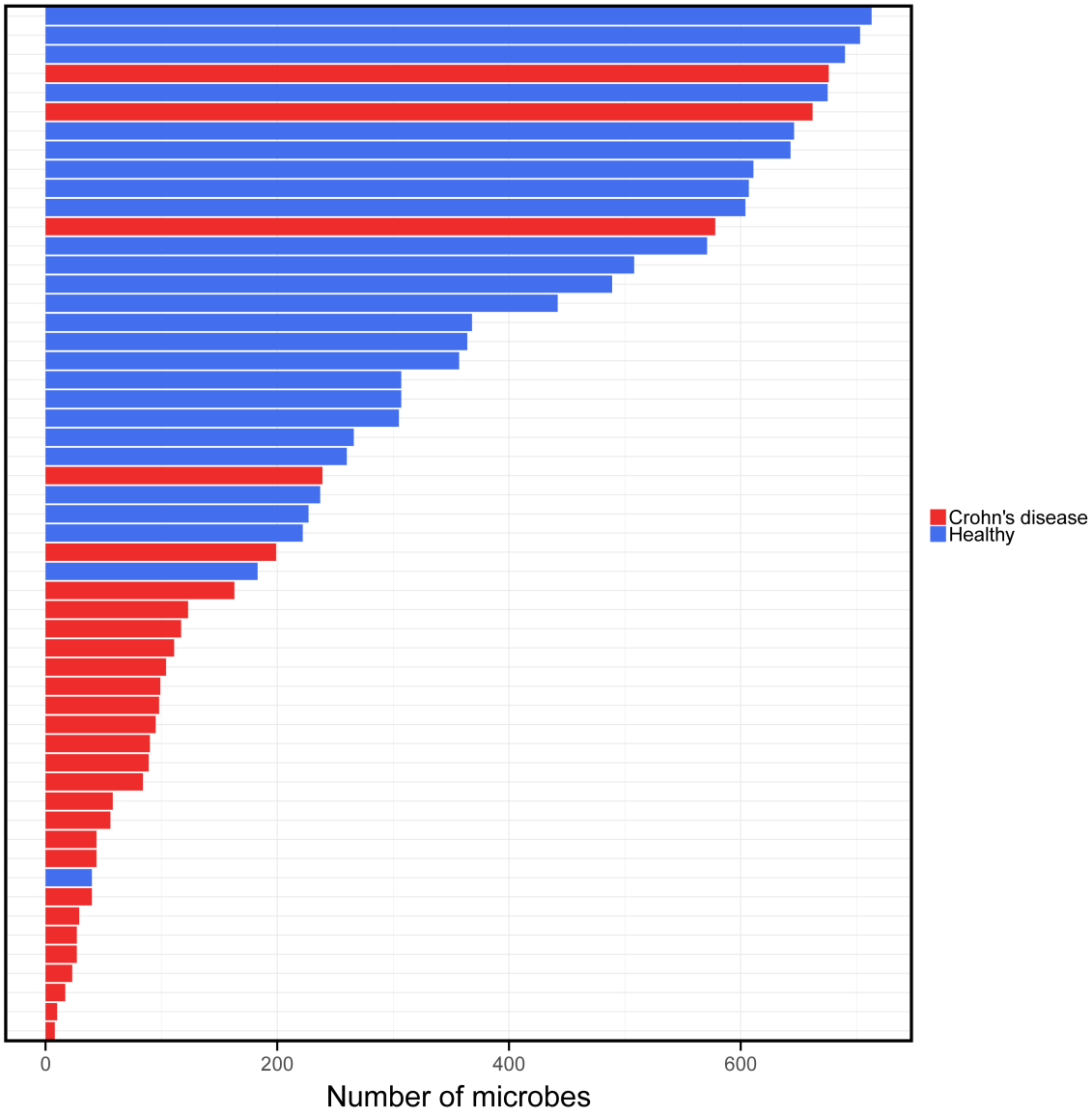

Figure C.1: Number of microbes that were detected to be present for each Crohn's disease patient and healthy control.

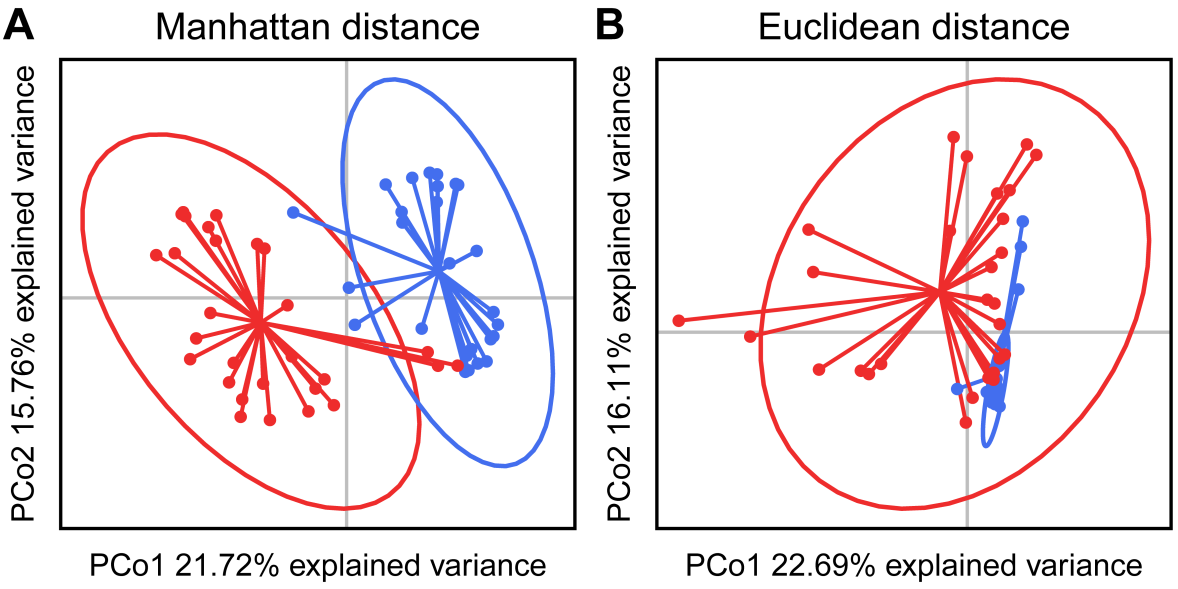

Figure C.2: Similarities between healthy controls and Crohn's disease patients assessed based on a principle coordinate analysis (PCoA) of the mapped abundance with Manhattan (A) and Euclidean distance (B).

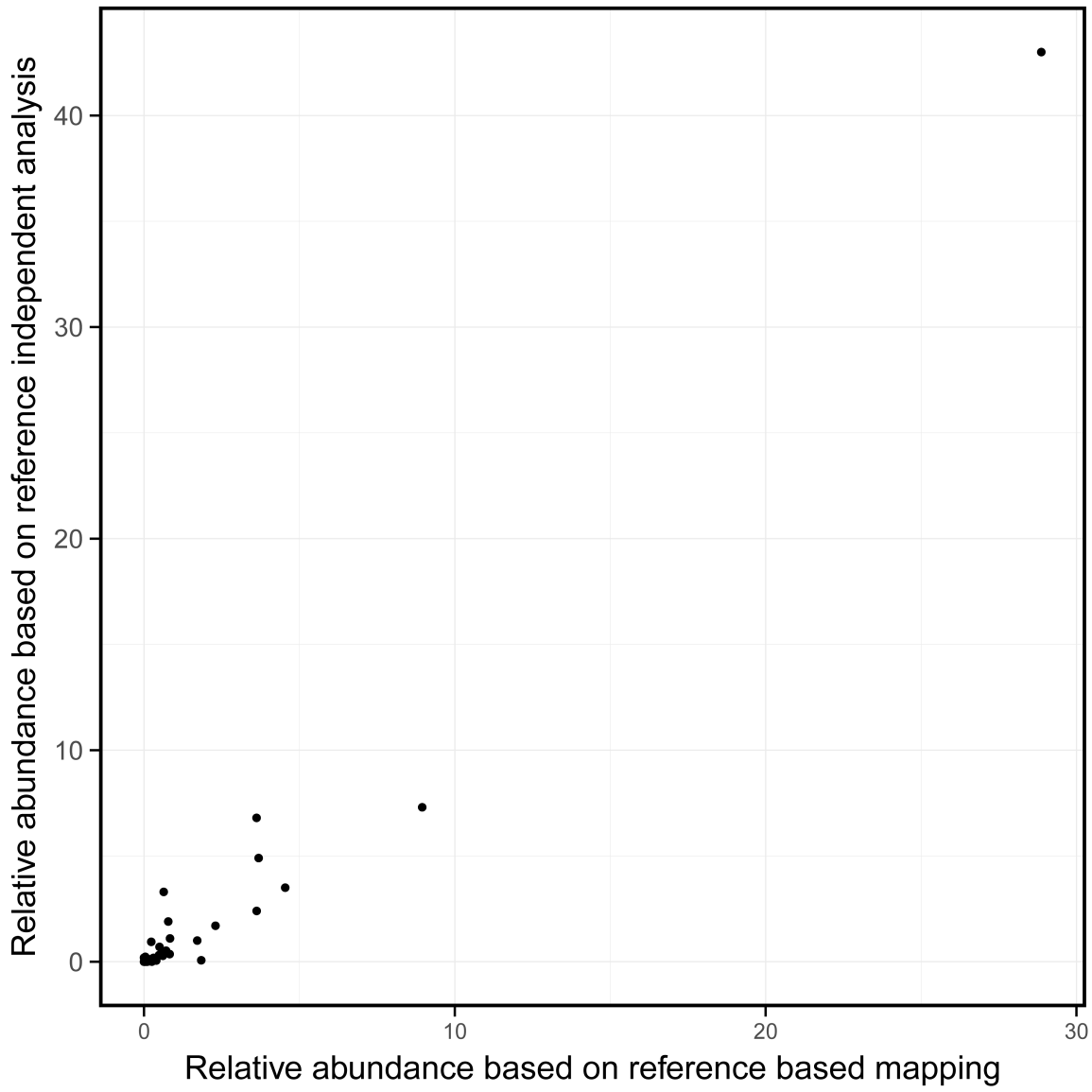

Figure C.3: Quantitative comparison of mapped microbe abundance values compared to the relative abundance of genera retrieved from the original study.

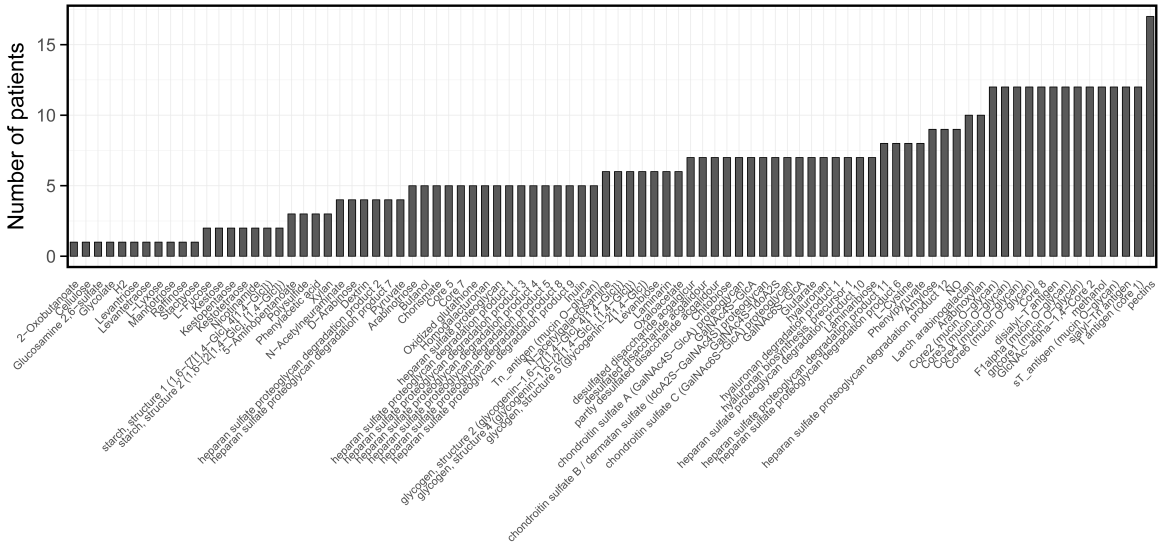

Figure C.4: Predicted metabolites used for in silico treatment of each patient. The number of patients is given in which the metabolite addition had a beneficial effect.