

Bioinformatics

doi.10.1093/bioinformatics/xxxxxx

Advance Access Publication Date: Day Month Year

Manuscript Category



Systems biology

# The Microbiome Modeling Toolbox: from microbial interactions to personalized microbial communities

Federico Baldini<sup>1</sup>, Almut Heinken<sup>1</sup>, Laurent Heirendt<sup>1</sup>, Stefania Magnúsdóttir<sup>1</sup>, Ronan M.T. Fleming<sup>2</sup>, and Ines Thiele<sup>1\*</sup>

<sup>1</sup>Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Campus Belval, Esch-sur-Alzette, Luxembourg and

<sup>2</sup>Division of Systems Biomedicine and Pharmacology, Leiden Academic Centre for Drug Research, Faculty of Science, Leiden University, Leiden, The Netherlands

\*To whom correspondence should be addressed.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

## Abstract

**Motivation:** The application of constraint-based modeling to functionally analyze metagenomic data has been limited so far, partially due to the absence of suitable toolboxes.

**Results:** To address this gap, we created a comprehensive toolbox to model i) microbe-microbe and host-microbe metabolic interactions, and ii) microbial communities using microbial genome-scale metabolic reconstructions and metagenomic data. The Microbiome Modeling Toolbox extends the functionality of the COBRA Toolbox.

**Availability:** The Microbiome Modeling Toolbox and the tutorials at <https://git.io/microbiomeModelingToolbox>.

**Contact:** [ines.thiele@gmail.com](mailto:ines.thiele@gmail.com)

## 1 Introduction

Microbial community sequencing data are increasingly available for numerous environmental niches (Mitchell, 2017). The analysis of this data often relies on investigating which microbes are present in a given sample. However, to further our understanding of the functional contribution of individual microbes in a community as well as the overall functional differences between communities, advanced analysis approaches, such as computational modeling, are required.

One possible approach is the constraint-based reconstruction and analysis (COBRA) approach, which builds genome-scale reconstructions of an organism and enables the prediction of, e.g., phenotypic properties (Palsson, 2006). Through the application of condition-specific constraints, an organism's metabolic reconstruction can be converted into many condition-specific models, which can be analyzed using available toolboxes, such as the Matlab (Mathworks, Inc.) based COBRA Toolbox (Heirendt *et al.*, 2017a). Metabolic reconstructions have been assembled for many organisms, including hundreds of gut microbes (Magnúsdóttir *et al.*, 2017) and human (Brunk *et al.*, 2018). While the COBRA Toolbox encapsulates many tools developed by the community for biotechnological and biomedical applications, it is currently focused on modeling single

organisms or cells. Here, we present the Microbiome Modeling Toolbox, which enables the generation, simulation, and interpretation of 1. pairwise microbe-microbe and host-microbe interactions, and 2. sample-specific microbial community models. By integrating sample-specific metagenomic data, the Microbiome Modeling Toolbox facilitates its analysis in the context of microbial reconstructions.

## 2 Features

**Pairwise interactions:** The pairwise interaction analysis determines metabolic exchange between two metabolic reconstructions. A joint matrix of two individual genome-scale reconstructions is generated, which enables them to freely exchange metabolites (Figure 1A). Defined nutrient input, e.g., a particular medium formulation, can be applied via the shared compartment using the corresponding exchange reactions. The pairwise microbial models can be investigated for six possible interaction types (i.e., competition, parasitism, amensalism, neutralism, commensalism, and mutualism) and Pareto optimality frontiers can be calculated. The tutorials *MicrobeMicrobeInteractions* and *HostMicrobeInteractions* illustrate the implemented functionalities.

**Microbial community modeling:** Metagenomic data can be analyzed using *mgPipe* (Figure 1B), which requires microbe identification and

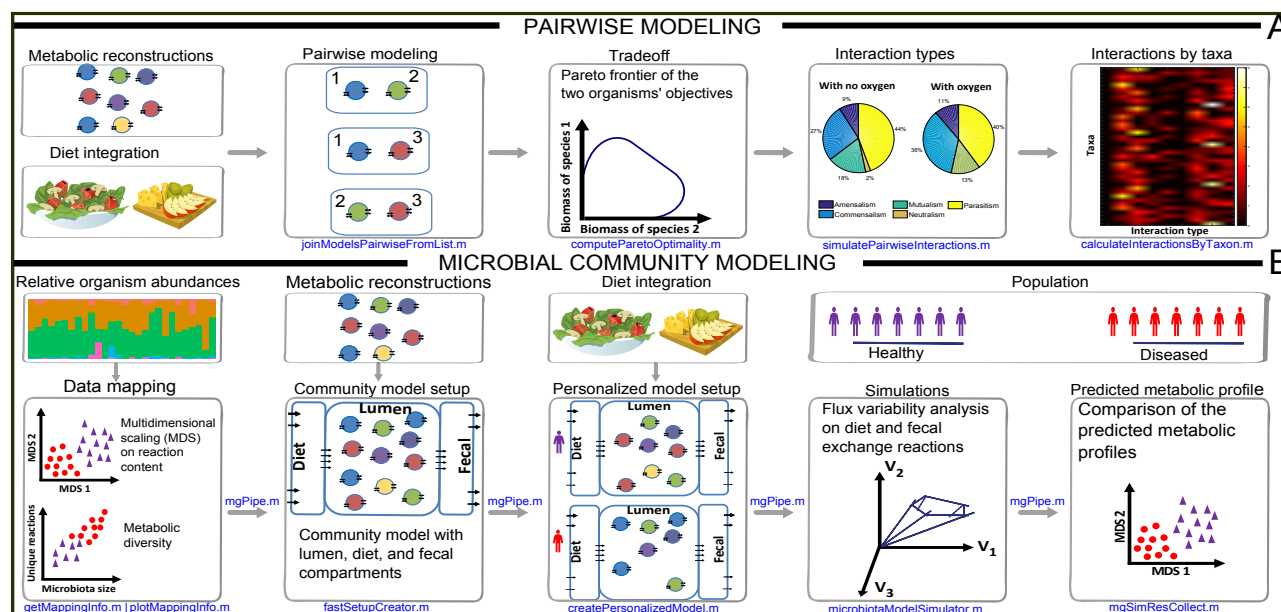


Fig. 1. Overview of the Microbiome Modeling Toolbox. A. Pairwise modeling of microbe-microbe and host-microbe interactions. B. Microbial community modeling.

relative abundance data for each sample, obtained with bioinformatic tools, such as QiIME 2 (Caporaso *et al.*, 2010) and MetaPhlAn (Segata *et al.*, 2012). *mgPipe* is divided into three parts: 1. the analysis of individuals' specific microbes abundances, including metabolic diversity and classical multidimensional scaling of the reactions in the identified microbes. 2. Construction of a personalized microbial community model using the identified microbes and their relative abundance data. For each personalized (or sample-specific) model, the corresponding microbial reconstructions are joined by adding reactions to each microbial reconstruction transporting metabolites from the extracellular space to the common lumen compartment. Metabolites present in the lumen compartment are connected to a diet and fecal compartment, enabling the uptake and secretion from/to the environment, respectively. Hundreds of reconstructions can be combined and modeled with using static parallelization. In each microbial community model, the community biomass reaction is personalized using the relative abundance data. Finally, coupling constraints (Heinken *et al.*, 2013) are applied to couple the flux through each microbial reaction to its corresponding biomass reaction flux. And 3. simulation of the personalized microbial community models under different diet regimes, e.g., using flux variability analysis (Heirendt *et al.*, 2017b). The differences between maximal uptake and secretion fluxes provide a metabolic profile for each microbial community sample, which can be analyzed using classical multidimensional scaling analyses. Diet-specific constraints (e.g., obtained from <https://vmh.life/#nutrition>) can be applied to the corresponding diet exchange reactions.

### 3 Implementation

The Microbiome Modeling Toolbox is written in MATLAB (Mathworks, Inc.) and accompanied with comprehensive documentation and tutorials. The toolbox allows for the integrative analysis of any number of reconstructions, including the human metabolic reconstruction (Brunk *et al.*, 2018). Metabolic reconstructions can be obtained from, e.g., the VHM (<https://vmh.life>), BioModels (<https://www.ebi.ac.uk/biomodels-main/>), and the KBase (<https://kbase.us/>). A uniform nomenclature of reaction and metabolite abbreviations across the reconstructions is

required. The implemented diet constraints require VMH abbreviations. To use higher taxonomical levels create pan-reconstructions (*createPanModels*). For larger datasets and/or bigger microbial community models, we recommend the use of the MATLAB command line or .m files and of a high-performance computing cluster.

### 4 Discussion

The Microbiome Modeling Toolbox enables the user to investigate microbial interactions at a large scale (Heinken *et al.*, 2013; Magnúsdóttir *et al.*, 2017). Moreover, metagenomically-derived data can be integrated with microbial metabolic reconstructions permitting the prediction of altered functional assessment of different microbial communities, e.g., in health and disease (Heinken *et al.*, 2018; Thiele *et al.*, 2018).

### Funding

This study received funding from the Luxembourg National Research Fund (FNR), through the ATTRACT programme (FNR/A12/01), and the OPEN grant (FNR/O16/11402054), as well as the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 757922).

### References

- Brunk, E., *et al.* (2018). Recon3d enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotech.*, **36**(3), 272.
- Caporaso, J. G., *et al.* (2010). Qiime allows analysis of high-throughput community sequencing data. *Nat. Methods*, **7**(5), 335.
- Heinken, A., *et al.* (2013). Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut microbes*, **4**(1), 28–40.
- Heinken, A. *et al.* (2018). Personalized modeling of the human gut microbiome reveals distinct bile acid deconjugation and biotransformation potential in healthy and IBD individuals. *bioRxiv*, doi.org/10.1101/229138.
- Heirendt, L., *et al.* (2017a). Creation and analysis of biochemical constraint-based models: the COBRA ToolBox v3. 0. *arXiv:1710.04038*.
- Heirendt, L., *et al.* (2017b). DistributedFBA. jl: High-level, high-performance flux balance analysis in julia. *Bioinformatics*, **33**(9), 1421–1423.

- Magnúsdóttir, S., *et al.* (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotech.*, **35**(1), 81.
- Mitchell, A. L. *et al.* (2017). EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucl Ac Res.*, **46**(D1): D726-D735.
- Palsson, B. (2006). Systems biology: properties of reconstructed networks. *Cambridge: Cambridge Univ Pr.*
- Segata, N. *et al.* (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Meth.*, **9**: 811-814.
- Thiele, I. *et al.* (2018). When metabolism meets physiology: Harvey and Harvetta. *bioRxiv*, doi.org/10.1101/255885.