Microbiome Modelling Toolbox 2.0: efficient, tractable modelling of microbiome communities

Almut Heinken1,2 and Ines Thiele1,2,3,4,*

1School of Medicine, National University of Galway, Galway, Ireland, 2Ryan Institute, National University of Galway, Galway, Ireland, 3Division of Microbiology, National University of Galway, Galway, Ireland and 4APC Microbiome Ireland, University College Cork, Cork, Ireland

*To whom correspondence should be addressed.

Abstract

Motivation: Constraint-Based Reconstruction and Analysis (COBRA) is a widely used approach for the interrogation and stratification of microbiome samples, yet applications to large-scale cohorts are hampered by limited scalability and efficiency of simulations.

Results: We substantially improved the computation speed and scalability of a previous implementation for the construction and interrogation of personalized constraint-based microbiome models as well as implemented additional functionalities for analysis and visualization.

Availability and implementation: Microbiome Modelling Toolbox and tutorials are freely available as part of the COBRA Toolbox at https://git.io/microbiomeModelingToolbox.

1 Introduction

The Constraint-Based Reconstruction and Analysis (COBRA) approach is a mechanistic, bottom-up systems biology method that relies on manually curated genome-scale reconstructions of metabolism that can be converted into mathematical models (Palsson, 2006). In recent years, COBRA has been used in >100 studies to interrogate the metabolism of microbial communities and microbiomes, yielding valuable insight into microbial community structure and function (Heinken et al., 2021). A notable application of COBRA modelling is the generation of personalized human microbiome models and their stratification based on their structure and function, which has been applied to, e.g., inflammatory bowel disease and colorectal cancer (Heinken et al., 2021).

We have previously developed the Microbiome Modelling Toolbox, a MATLAB-based suite for the construction and interrogation of microbe–microbe and personalized microbiome models (Baldini et al., 2019). Here, we present an updated version, the Microbiome Modelling Toolbox 2.0. Compared with its predecessor, parallelization has been optimized in both modules for improved computation speed. The mgPipe module has been expanded by additional functions, resulting in a comprehensive analysis pipeline as described below.

2 Features

2.1 Generation of personalized microbiome models

mgPipe enables the generation of personalized microbiome models from, e.g., a large collection of microbial metabolic reconstructions, aka AGORA (Heinken et al., 2020), using the microbial composition information from either 16s rRNA or metagenomic sequencing data as input to build, interrogate and analyse personalized microbiome models, which allows for the stratification of, e.g., disease cases and controls according to their microbiomes’ metabolic potential (Baldini et al., 2019). Compared with its predecessor, parallelization has been optimized in both modules for improved computation speed. The mgPipe module has been expanded by additional functions, resulting in a comprehensive analysis pipeline as described below.

2.2 Computation of microbiome-level metabolic fluxes

mgPipe provides functions for both comprehensive and targeted simulation of microbiome metabolic fluxes, corresponding to the prediction of interactions, such as mutualism, commensalism and competition (Baldini et al., 2019). Second, the mgPipe module uses microbial composition information obtained from 16s rRNA or metagenomic sequencing data as input to build, interrogate and analyse personalized microbiome models, which allows for the stratification of, e.g., disease cases and controls according to their microbiomes’ metabolic potential (Baldini et al., 2019). Compared with its predecessor, parallelization has been optimized in both modules for improved computation speed. The mgPipe module has been expanded by additional functions, resulting in a comprehensive analysis pipeline as described below.

© The Author(s) 2022. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
predicted fecal metabolome (Fig. 1). To comprehensively profile each microbiome model, every metabolite, which can be transported by at least one microbe, is routinely computed through computationally efficient flux variability analysis (Gudmundsson and Thiele, 2010). Metabolite fluxes can be subsequently correlated with taxon abundances. As a computationally efficient option, mgPipe also allows for the computation of selected metabolites while also retrieving shadow prices, a measurement for the value of a metabolite towards the objective function (Palsson, 2006). To directly identify microbe-metabolite links, microbes contributing most to the uptake and secretion of metabolites can also be retrieved. Finally, structural features of the microbiome models, such as dimension, reaction abundance, and subsystem abundance, are computed (Fig. 1).

2.3 Visualization and further analysis
Compared with its predecessor, the Microbiome Modelling Toolbox 2.0 provides additional functions for visualization and stratification of fluxes based on available metadata (e.g., disease state). Principal components analysis as well as detailed plots for each computed metabolite stratified by group can be generated (Fig. 1). mgPipe also provides functions for determining whether differences in model properties and metabolite production flux between stratification groups are significant (Fig. 1).

3 Implementation
The Microbiome Modelling Toolbox 2.0 is written in MATLAB (Mathworks, Inc.) and is freely available at the COBRA Toolbox GitHub repository, https://github.com/opencobra/cobratoolbox (Heirendt et al., 2019). Comprehensive tutorials in form of MATLAB live scripts are provided at https://github.com/opencobra/COBRA.tutorials. Further details are provided at https://git.io/microbiomeModelingToolbox.

4 Discussion
Recent cohort studies have generated an unprecedented amount of publicly available metagenomic sequencing data of human microorganisms varying, e.g., in health status, diet, and ethnicity. Yet, the interpretation of this wealth of data is lagging. The increasing availability of strain-resolved metagenomic data, as well as of curated genome-scale reconstructions, such as the AGORA resource of 773 reconstructed human gut microbes (Magnusdottir et al., 2017), and its expansion, AGORA2, now containing over 7,000 microbial reconstructions (Heinken et al., 2020), allow for the construction of personalized microbiome models with increasing coverage and hence, size. The Microbiome Modelling Toolbox 2.0, due to its improved scalability and efficiency, provides the computational tools for the large-scale interrogation of hundreds (Hertel et al., 2021) or even thousands of microbiomes. For instance, it enabled the comprehensive prediction of the fecal metabolome for 644 microbiome models of colorectal cancer patients and controls containing up to ~150 species and ~240,000 reactions (Hertel et al., 2021), as well as the construction of over 14,000 personalized microbiome models (A.Heinken et al., in preparation). Personalized models generated by mgPipe can also be integrated with a whole-body model of human metabolism (Thiele et al., 2020), hence, enabling constraint-based modelling of human–microbiome interactions with unprecedented scope.

Funding
This work was supported by grants from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme [757922] to I.T. and the National Institute on Aging [1RF1AG058942-01 and 1U19AG063744-01].

Conflict of Interest: none declared.

References
Baldini,F. et al. (2019) The Microbiome Modeling Toolbox: from microbial interactions to personalized microbial communities. Bioinformatics, 35, 2332–2334.
Gudmundsson,S. and Thiele,I. (2010) Computationally efficient flux variability analysis. BMC Bioinform., 11, 489.
Heinken,A. et al. (2020) AGORA2: large scale reconstruction of the microbiome highlights wide-spread drug-metabolising capacities. bioRxiv. 2020.2011.2009.375451.
Heinken,A. et al. (2021) Advances in constraint-based modelling of microbial communities. Curr. Opin. Syst. Biol., 27, 100346.
Heirendt,L. et al. (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat. Protoc., 14, 639–702.
Hertel,J. et al. (2021) Integration of constraint-based modeling with fecal metabolomics reveals large deleterious effects of Fusobacterium spp. On community butyrate production. Gut Microbes, 13, 1–23.
Magnusdottir,S. et al. (2017) Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat. Biotechnol., 35, 81–89.
Palsson,B. (2006) Systems Biology: Properties of Reconstructed Networks. Cambridge University Press, Cambridge, New York.
Thiele,I. et al. (2020) Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. Mol. Syst. Biol., 16, e8982.