



HAL
open science

Community-scale models of microbiomes: articulating metabolic modelling and metagenome sequencing

Klara Cerk, Pablo Ugalde-Salas, Chabname Ghassemi Nedjad, Maxime Lecomte, Coralie Muller, David James Sherman, Falk Hildebrand, Simon Labarthe, Clémence Frioux

► To cite this version:

Klara Cerk, Pablo Ugalde-Salas, Chabname Ghassemi Nedjad, Maxime Lecomte, Coralie Muller, et al.. Community-scale models of microbiomes: articulating metabolic modelling and metagenome sequencing. *Microbial Biotechnology*, 2024, 10.1111/1751-7915.14396 . hal-04409251

HAL Id: hal-04409251

<https://inria.hal.science/hal-04409251v1>

Submitted on 22 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Community-scale models of microbiomes: articulating metabolic modelling and metagenome sequencing

Klara Cerk^{1,2}, Pablo Ugalde-Salas⁴, Chabname Ghassemi Nedjad^{3,4}, Maxime Lecomte^{5,4}, Coralie Muller⁴, David J. Sherman⁴, Falk Hildebrand^{1,2}, Simon Labarthe^{6,4}, Clémence Frioux⁴

¹ *Quadram Institute Bioscience, NR4 7GJ Norwich, United Kingdom*

² *Earlham Institute, NR4 7GJ Norwich, United Kingdom*

³ *Univ. Bordeaux, Bordeaux INP, CNRS, LaBRI, F-33405 Talence, France*

⁴ *Inria, Univ. Bordeaux, INRAE, F-33405 Talence, France*

⁵ *Univ. Rennes, INRAE, STLO, F-35042 Rennes, France*

⁶ *Univ. Bordeaux, INRAE, BIOGECO, F-33610 Cestas, France*

Correspondence: Clémence Frioux (ORCID: 0000-0003-2114-0697) <clemence.frioux@inria.fr>

CC-BY 4.0

RUNNING TITLE. Metagenomics and metabolic modelling in microbiomes

ABSTRACT. Building models is essential for understanding the functions and dynamics of microbial communities. Metabolic models built on genome-scale metabolic network reconstructions (GENREs) are especially relevant as a means to decipher the complex interactions occurring among species.

Model reconstruction increasingly relies on metagenomics, which permits direct characterisation of naturally-occurring communities that may contain organisms that cannot be isolated or cultured. In this review, we provide an overview of the field of metabolic modelling and its increasing reliance on and synergy with metagenomics and bioinformatics. We survey the means of assigning functions and reconstructing metabolic networks from (meta-)genomes, and present the variety and mathematical fundamentals of metabolic models that foster the understanding of microbial dynamics. We emphasise the characterisation of interactions and the scaling of model construction to large communities, two important bottlenecks in the applicability of these models.

We give an overview of the current state of the art in metagenome sequencing and bioinformatics analysis, focusing on the reconstruction of genomes in microbial communities. Metagenomics benefits tremendously from third generation sequencing, and we discuss the opportunities of long read sequencing, strain-level characterisation, and eukaryotic metagenomics. We aim at providing algorithmic and mathematical support, together with tool and application resources, that permit bridging the gap between metagenomics and metabolic modelling.

KEYWORDS. Metagenomics - Metabolic networks - Modelling - Microbial community - Third generation sequencing

1. Introduction

The roles of microbiomes in the environment and in health-associated, agricultural or biotechnological applications is widely acknowledged [Peixoto et al., 2022, Foundation et al., 2022, Wilmes et al., 2022]. These assemblies of microorganisms are governed by the interactions they have with their environment, and by those occurring among species [Zengler and Zaramela, 2018]. Such ecological interactions can generally be classified as neutral, positive, or negative; and can be inferred through network analysis [Faust and Raes, 2012]. Mechanistically, they operate at a similar level, called the *metabolism*. Metabolism encompasses all of the functions of cells, defined by biochemical reactions and by molecular exchanges between cells and the environment. Microbiomes are characterised by their biosynthetic potential and by the taxonomic diversity that they harbour [Heintz-Buschart and Wilmes, 2018]. While some microbiomes such as the gastro-intestinal tract in humans have been extensively studied, thereby revealing insights into the putative cooperation and competition networks among bacteria [Coyte and Rakoff-Nahoum, 2019], in general the precise mechanisms, dynamics and control levers of metabolic interactions remain poorly understood. These difficulties are even more prevalent in complex environmental microbiomes where the massive diversity of micro-organisms and their poor representation in public databases impede taxonomic and functional surveys [Brown et al., 2021].

Culture-independent methods can be used to broadly characterise microbiomes [Zhang et al., 2019].

55 Prominent among these is metagenomics, the random sequencing of all DNA in a given sample. Metage-
56 nomics is by now ubiquitously applied to decipher both the taxonomic and functional diversity of samples;
57 a rich resource of public available metagenomes can support large scale studies available to even small
58 labs with the required bioinformatics expertise [Gurbich et al., 2023]. Analysis of metagenomic data by
59 bioinformatic tools can generate surveys of the functions embodied by the community [Kindler et al., 2021].
60 Metabolic models, either mechanistic models such as those reconstructed from genome-scale metabolic net-
61 work reconstructions (GENREs), or phenomenological models describing phenotypic traits, go further than
62 these surveys. The latter have been the modeller’s choice for narrowing down the complexity of community-
63 wide behaviour [van den Berg et al., 2022], whereas the former could originally only be applied to single
64 species [Varma and Palsson, 1994]. However the range of modelling tools has strongly increased during the
65 past years, and now provides a broad collection of methods for answering critical questions about microbial
66 communities [Kumar et al., 2019]: what are the roles of bacterial species in the ecosystem? What functions
67 are carried and how redundant are they? How will the community behave over time? And, can we alter
68 this behaviour, or modify the community composition to reach a desired state?

69 Recent developments in metabolic model inference facilitate this procedure and make it worthwhile
70 to consider systematic reconstruction for collections of genomes [Frioux et al., 2020]. Concurrently, al-
71 gorithmic resources for obtaining such genomes from metagenomes have increased, and the technological
72 development of third generation sequencing (TGS) has paved the way to a major change of paradigm in the
73 study of microbiomes [Sereika et al., 2022]. Our goal here is twofold: (1) to review to what extent metabolic
74 modelling and metagenomics can interact in service of better comprehending microbial ecosystems, (2) to
75 provide comprehensive lists of practical tools that can be used by non-specialists for metagenomics assem-
76 bly, metabolic reconstruction, and microbiome modelling. We describe the different kinds of models that
77 can be built from metagenomic data to predict the functions and the dynamics of a microbial community,
78 and how they can be used to infer interactions among species. As these models rely on reconstructions of
79 metagenomes, we present as well the latest sequencing techniques, bioinformatics tools and open challenges
80 related to metagenomics, in an attempt to draw together these two domains of computational biology.

81 **2. From gene functions to metabolic reconstructions**

82 **2.1 Aggregating functions into metabolic networks**

83 Models of microbial communities can be derived from a subset of functions, reactions, or metabolic path-
84 ways; or on the contrary by analysing the entire metabolic potential of the genomes. In the latter case,
85 the exploration of the link between the genome, enzymes, and the metabolic reactions they catalyse
86 provides a system-wide understanding of the physiology of the cell and its links with its environment.
87 *Genome-scale metabolic network reconstructions* (GENREs) contain metabolic reactions predicted from
88 the entire genomic content through gene-protein-reaction (GPR) relationships, but also spontaneous reac-
89 tions and transport reactions connecting the cell to its environment or even to intracellular components
90 [Thiele and Palsson, 2010]. In this section, we present the basis of such metabolic networks, whose fun-
91 damental concept is the association of genes to functions. Since GENREs are at the heart of mechanistic
92 models of metabolism, the first step of community metabolic modelling is typically the acquisition of a
93 GENRE for each of its members.

94 GENREs have been built for more than twenty years, with a spectacular acceleration of the publications
95 over the past years, as a result of both the increased availability of genomes, and of the efforts of the
96 community for tools development. In 2019, a review tallied more than 6,000 published GENREs, 94% of
97 which were for bacteria [Gu et al., 2019]. However, the numbers rise even more when including systematic
98 and automatic reconstructions performed on publicly available genomes by platforms such as BioCyc
99 [Caspi et al., 2019], or AGORA2 for human-associated micro-organisms [Heinken et al., 2023]. Table 1
100 gathers databases that can be used as general or specialised resources to functionally annotate genomes
101 prior to GENRE generation, as well as collections of metabolic reconstructions. In addition, *de novo*
102 reconstruction is facilitated by a variety of software programmes and toolboxes created for high throughput
103 processing of genomes, an important criterion when tackling large-scale communities of micro-organisms.
104 While many of tools, catalogued in dedicated reviews [Mendoza et al., 2019], enable this task, only a small
105 number can do so fully automatically, without user interaction. We hereafter present a sample of tools
106 that we believe are suitable for systematic metabolic reconstruction of GENREs starting from a list of
107 annotated genomes.

108 Gapseq [Zimmermann et al., 2021] builds GENREs using a curated reaction database and a dedicated
109 gap-filling algorithm. The tool was notably applied to a four-strain microbial community whose associated

Table 1. Main generalist and specialised functional annotation databases and resources of existing metabolic network reconstructions

Category	Name	Content	Application for metabolic modelling
Generalist databases	KEGG	Several databases including biochemical reactions [Kanehisa et al., 2021]	General annotation of genes and genomes that can be directly or indirectly (via reconstruction tools) associated to metabolic reactions
	COG	Clusters of orthologous genes [Galperin et al., 2020]	
	Gene Ontology	Knowledge base of gene functions [Galperin et al., 2020]	
	UniProt	Association of proteins to functions [Consortium et al., 2021]	
	Pfam	Classification of proteins into families and domains [Mistry et al., 2020]	
	BRENDA	Collection of enzymes and their metabolic reactions [Chang et al., 2020]	
Metabolic reconstruction collections	eggNOG	Orthologous groups and functional annotation [Huerta-Cepas et al., 2018]	Ready-to-use reconstructions and models
	MetaCyc	Encyclopaedia of metabolic reactions [Caspi et al., 2019]	
	BioCyc	More than 20,000 pathway-genome databases [Karp et al., 2019a]	
	BiGG	High quality metabolic network reconstructions [King et al., 2016]	
	BioModels	Literature -associated (metabolic) models [Malik-Sheriff et al., 2019]	
Specialised databases	AGORA2	GENREs of 7,302 human-associates micr-organisms [Heinken et al., 2023]	Curation, validation, inspection of model predictions
	EMBL GEMS	GENREs reconstructed with Carveme for 5587 reference and representative bacterial genomes of NCBI RefSeq [Machado et al., 2018]	
	CAZy	Carbohydrate-active enzymes [Drula et al., 2022]	
	CARD	Antibiotic resistance [Alcock et al., 2020]	
	TCDB	Transporter classification [Saier et al., 2021]	
	Norine	Non-ribosomal peptides [Flissi et al., 2020]	
	antiSMASH	Natural products of the secondary metabolism [Blin et al., 2020]	
	MEROPS	Proteolytic enzymes and their substrates [Rawlings et al., 2017]	

110 models reproduced expected behaviours of the gut microbiome metabolism. Carveme [Machado et al., 2018]
111 uses a top down approach consisting in pruning (carving) a universal curated model of the microbial
112 metabolism for rapid generation of GENREs starting from genomes. While construction with Gapseq
113 takes up to two hours, a Carveme-generated reconstruction can be obtained in a few minutes. Both
114 computation times are compatible with a systematic generation of GENREs for microbiomes. Pathway
115 Tools [Karp et al., 2019b] builds pathway-genome databases relying on the MetaCyc database of genomes
116 [Caspi et al., 2019] and on functionally annotations, e.g. from EggNOG-mapper [Cantalapiedra et al., 2021],
117 Prokka [Seemann, 2014] or RAST [Overbeek et al., 2014]. While Pathway Tools is originally based on an in-
118 teractive single-organism interface, which can impede its use at the scale of large communities, it is possible
119 to automatise the process using, for instance, the Python package mpwt [Belcour et al., 2020]. The latter
120 was applied to a collection of 336 reference genomes obtained after metagenomic read mapping, in order to
121 identify the presence of pathways of interest [Ruuskanen et al., 2021]. The latest developments of Pathway
122 Tools additionally permit importing a set of genomes or metagenome-assembled genomes (MAGs) (see sec-
123 tion 7) into the interface to run the reconstruction on each organism. AutoKEGGRec [Karlsen et al., 2018]
124 is a Matlab programme that creates draft GENREs by retrieving reactions and GPR relationships from
125 the KEGG database for one or a set of organisms, making it suitable for community modelling. In addi-
126 tion, online platforms provide reconstruction services, such as KBase [Arkin et al., 2018], which proposes a
127 large variety of analyses including metabolic reconstruction, with extensive efforts towards reproducibility

128 through the use of narratives; or ModelSEED [Seaver et al., 2020], which performs metabolic annotation
129 and GENRE reconstruction for microbes, plants and fungi.

130 A general limitation of GENREs lies in the metabolic landscape covered in reaction databases; many
131 proteins have no attributed functions even in well-studied organisms, and experimental evidence of func-
132 tions are even scarcer [Monk et al., 2014]. It must be noted that automatically reconstructed GENREs
133 cannot compete with manually curated ones with respect to the quality of the contents and the accuracy of
134 the simulations they permit. High-quality GENREs exist mostly for model organisms; automatically recon-
135 structed GENREs may suffer from erroneous GPR relationships or reversibility constraints, and inaccuracy
136 in biomass composition [Gu et al., 2019]. Most automated tools produce GENREs that are gap-filled in
137 order for the associated model to simulate the production of biomass [Feist and Palsson, 2010], which
138 is a common step for flux prediction (see section 3.1). Unfortunately, automated gap-filling, while use-
139 ful, is not as reliable as manual curation for producing high quality GENREs [Latendresse and Karp, 2018,
140 Karp et al., 2018]. Additionally, growth is simulated in media that likely do not sustain the auxotrophies of
141 non-cultivable organisms, which may depend on complex metabolic interactions with other microbes to grow
142 [West et al., 2007, Zengler and Zaramela, 2018]. On the other hand, it is possible to work with non-gap-
143 filled GENREs, for example using Pathway Tools or AutoKEGGRec. Overall, the use of simulation-ready
144 GENREs creates a risk of missing interactions occurring in communities (false negative interaction) while
145 using non-gap-filled GENREs may falsely predict interactions to sustain certain functions that could actu-
146 ally be performed by individual microbes (false positive interaction). We refer to [Bernstein et al., 2021] for
147 a description of all the steps of the GENRE reconstruction and simulation processes where uncertainty may
148 occur and lead to inconsistent results. More generally, with the increasing use of metagenomic-assembled
149 genomes (see section 7), more attention is paid to the effect of genome incompleteness on the inference of
150 functions and metabolic potentials [Eisenhofer et al., 2023, Belcour et al., 2020]. The examination of mul-
151 tiple GENREs when building models of microbiomes must therefore consider these potential shortcomings,
152 and address them through a careful choice of model paradigm and a cautious interpretation of results.

153 Obtaining a GENRE for each member of a microbial community is already *per se* a valuable resource for
154 querying and comparing the metabolic processes that can occur within species, even before modelling and
155 simulation. It can consist in identifying functional differences between GENREs: strain-specific patterns,
156 gene essentiality or biosynthetic capacity [Biggs et al., 2015]. In [Weiss et al., 2022], GENREs correspond-
157 ing to the twelve bacterial members of a synthetic bacterial community were compared at the level of

158 metabolic pathway, as a complement to results of *in vitro* interaction experiments. More generally, any
159 approach aiming at comparing metabolic networks could be applied to GENREs of co-occurring species
160 [Ay et al., 2012]. Likewise, studying the pan-reactome and accessory reactome of species using GENREs
161 is valuable [Prigent et al., 2018, Belcour et al., 2023] and these approaches can be applied to ecosystems.

162 **3. Metabolic models for microbial communities: from GENREs to** 163 **dynamic simulations**

164 GENREs can be the input of metabolic modelling at the genome or community scale. In this section, we
165 describe and focus on mechanistic constraint-based models. Alternative GENRE-based models and their
166 relationship to other modelling fields are presented in section 5.

167 When modelling a community, the first step is defining a collection of individual GENREs representative
168 of the community. The selection can be ‘reductionist’, meaning a reduced number of genomes is selected
169 as functionally representative of the community and a GENRE is obtained for each, either from *de novo*
170 reconstruction, or from databases (Table1). A second ‘exhaustive’ approach consists in building a GENRE
171 for each genome encountered in the community e.g. from metagenomic analysis. Then, two alternative
172 conceptual views exist to build a community model, with methodological and computational pros and cons
173 (Fig. 1): one consisting in merging GENREs to build a *a community-wide model*, also termed *bag-of-genes*
174 *model* and the other articulating individual GENREs into a *community of GENREs (bag-of-genomes)*
175 [Frioux et al., 2020].

176 **3.1 Genome-scale metabolic modelling**

Metabolic networks, metabolic models and dynamic simulations A GENRE has a matrix rep-
resentation termed the *stoichiometric matrix* \mathcal{S} , with as many rows as metabolites and as many columns
as reactions (Fig. 1). The coefficient \mathcal{S}_{ij} is the stoichiometric coefficient of the metabolite i in the reaction
 j , with a negative or positive value indicating whether i is a substrate or a product of j , or a zero value
if the metabolite i does not appear in reaction j [Orth et al., 2010]. The matrix \mathcal{S} allows modelling the
dynamics of the metabolite concentrations c (mmol L^{-1}) b (g L^{-1}) through the mass-balance equation

$$\partial_t c = \mathcal{S}v(c, b) \quad (1)$$

177 where ∂_t is the time derivative operator and $\nu(c, b)$ ($\text{mmol h}^{-1} \text{L}$) is a vector of the reaction fluxes that
178 depends both on c and on the bacterial population b [Jamshidi and Palsson, 2010]. Equation (1) highlights
179 that modelling concentration variation comprises two distinct components: (1) the metabolic network,
180 i.e. the static functional potential carried by the micro-organism genome represented by \mathcal{S} , and (2) the
181 distribution of the metabolic fluxes on this network, i.e. the activated functions in the current state of the
182 system, represented by ν . The resulting model is referred to as a *genome-scale metabolic model* (GEM).
183 The algebraic analysis of \mathcal{S} already gives biological insights into the metabolic graph, through the use of
184 approaches such as elementary flux modes (EFM) [Stelling et al., 2002].

185 **Constraint-based models for flux prediction** To predict ν for a single species or a merged com-
186 munity (see fig. 1), several approaches are available. *Kinetic modelling* adds the enzymes catalysing
187 the reactions to the system and evaluates ν using Michaelis-Menten-like equations that link substrate
188 and enzyme concentrations to the reaction kinetics [Costa et al., 2016]. However, due to the difficulty of
189 model parametrisation, which requires enzyme kinetic rates, practical use is limited at a genome scale
190 [van Rosmalen et al., 2021] and even more so at metagenome scale. *Constraint-based metabolic models* are
191 a widely used alternative for predicting flux distributions, that assumes a quasi steady-state of the intra-
192 cellular metabolites in eq. 1. *Flux Balance Analysis* (FBA) [Orth et al., 2010] solves a linear optimisation of
193 an objective function—usually the biomass production—and taking into account thermodynamic constraints
194 modelled as flux bounds.

195 Known limitations of FBA are the need for high-quality or curated GENREs as input, and its sensitivity
196 to the definition of the biomass function, which is organism-dependent [Xavier et al., 2017] and is hard to
197 establish for a community-wide model. Another limitation is the existence of many alternative optimal so-
198 lutions to the problem. Numerous variations exist, such as Parsimonious FBA (pFBA) [Lewis et al., 2010],
199 which focuses on sparse flux distributions; Flux Variability Analysis (FVA) [Mahadevan and Schilling, 2003],
200 which explores the range of alternative solutions to identify essential and blocked reactions; Thermody-
201 namic Flux Analysis (TFA) [Henry et al., 2007], which identifies thermodynamically feasible flux distribu-
202 tion; dynamic FBA (dFBA) [Mahadevan et al., 2002], which iterates flux evaluations and concentration
203 updates to model the dynamics of the system; and Resource Balance Analysis (RBA) [Goelzer et al., 2015],
204 which extends the optimisation problem to take into account the internal machinery of protein produc-
205 tion. We refer to [León and Nogales, 2022] for an extensive review of constraint-based methods discussing
206 the biological trade-offs they achieve. The Cobra toolbox [Heirendt et al., 2019] and its Python interface

207 [Ebrahim et al., 2013] provide numerous practical tools and implementations for constraint-based mod-
208 elling.

209 Several methods integrate -omics data into constraint-based models, such as transcriptomic [Jenior et al., 2020]
210 or metabolomic data [Toya and Shimizu, 2013]. Metatranscriptomics data can support the creation of
211 condition-specific community metabolic models such as in the work of [Zampieri et al., 2023]. Going be-
212 yond the analysis of functional potential—which may not be expressed—this approach takes into account
213 whole transcriptome abundance of MAGs and gene transcript abundance, to scale and adapt metabolic
214 reaction bounds. Such integrated solutions bring metabolic simulations closer to the biological reality.

215 Building constraint-based models using GENREs of individual microbiome members is a first step in
216 the comparison of their biosynthetic potential [Bartell et al., 2014]. As one approach, CONGA compares
217 GENREs pairwise to identify essential genes [Hamilton and Reed, 2012]. Pangenome analysis combined
218 with flux analysis can also be used to detect strain specific metabolic features [Bosi et al., 2016] (see also
219 section 6.3). A further step in the analysis is to combine individual GEMs in order to model the community
220 and its interactions.

221 **3.2 From genome to metagenome-scale metabolic modelling**

222 **Bag-of-genomes models** A batch of GENREs can first be considered as a set of independent individuals
223 interacting through extracellular metabolites. The community model is then obtained by the juxtaposition
224 of the individual GENREs, which are used to predict individual metabolic fluxes, the sum of which deter-
225 mines the fate of the extracellular metabolites (Fig. 1). This paradigm is also termed a *bag-of-genomes*
226 model [Frioux et al., 2020].

227 However, the prediction of individual fluxes can require new methodological developments, in particular
228 in a multi-species FBA framework where interactions are studied in a static regime [Gottstein et al., 2016].
229 Indeed, the growth rate of an individual only depends on the availability of extracellular metabolites,
230 whereas that of a community constraint-based model also depends on the instantaneous use of metabolites
231 by organisms that limits metabolic availability for others. Individual flux prediction in communities is
232 thus the result of a trade-off of metabolite uptake between the community members. Furthermore, the
233 hypothesis of individual optimal growth, which is sound in monoculture, may not be valid in communities
234 where complex ecological processes may occur. For example, symbiosis can induce sub-optimal growth for
235 a symbiont but improve long-term persistence in the community. To predict this trade-off, several meth-

Table 2. Main methods and corresponding framework to solve community FBA and dFBA models

Framework	Method	Description	platform and references
FBA	multi-level optimisation	An inner optimisation is made for each individual GENRE and coupled to an outer community-wide optimisation: trade-off between individual and total biomass growth	Optcom[Zomorodi and Maranas, 2012] CASINO[Shoae et al., 2015]
	weighted squared sum	A weighted square sum of individual flux weighted by microbe abundance as community-scale objective function	MICOM [Diener et al., 2020]
	multi-objective FBA	A Pareto front is computed as set of solutions of a multi-objective optimisation problem. A solution on the Pareto front is chosen according to additional criteria (e.g. resource use)	MO-FBA [Budnich et al., 2017] [Heinken and Thiele, 2015] [Heinken et al., 2013]
	EFM-based	EFM-based individual models the kinetic rates of which is determined by a community-scale model	CODY[Geng et al., 2021]
dFBA	sum of fluxes	Straightforward sum of individual fluxes (see fig.1). Flux bounds are computed from extracellular compound concentrations that change over time according to consumption and production by community members.	μ BialSim [Popp and Centler, 2020] COMETS [Dukovski et al., 2021] IndiMeSH [Borer et al., 2019]

ods have been developed, that are conceptually very different in a static or a dynamic FBA framework. These methods and example of corresponding toolboxes are recapitulated in table 2. Extensive reviews of community-wide metabolic models can be found in [Colarusso et al., 2021], including comparison between FBA and EFM-based methods, in [León and Nogales, 2022], for a comparison between top-down and bottom-up modelling approaches, or in [García-Jiménez et al., 2021, Zaramela et al., 2021], for a parallel between computational methods and engineered microbial communities.

The Microbiome Modelling Toolbox [Heinken and Thiele, 2022a] derives community models from metagenomics data for the human gut microbiome. This platform facilitates the extraction of GEMs corresponding to OTUs identified by metabarcoding from large curated collections such as AGORA 2 [Heinken et al., 2023] or the Virtual Metabolic Human [Noronha et al., 2018], and provides efficient computation and result visualisation. The main drawbacks of such large bag-of-genomes models are the computational cost, which scales with the number of GEMs taken into account [Heinken and Thiele, 2022a], and the difficulty of ensuring GENRE quality and accuracy for a large number of genomes. In particular, the presence of draft models inferred by automatic pipelines may impair the overall quality of the simulations [Babaei et al., 2018]. Bag-of-genomes models can be embedded in an outer optimisation problem in order to find an optimal configuration of the microbial community according to a dedicated fitness function, like in FLYCOP [García-Jiménez et al., 2018].

Bag-of-genes models A community-wide metabolic model can be built by assembling all the functions carried by the individual GENREs and factoring the common functions shared by different organisms (Fig. 1). In this paradigm, the community is seen as a meta or supra organism, an *enzyme soup*, whose functional potential results from the gathering of the whole set of functions [Frioux et al., 2020]. The corresponding

257 stoichiometric matrix \mathcal{S}_{com} is the concatenation of the individual stoichiometric matrices after removal of
258 duplicated metabolites and functions.

259 Bag-of-genes models have been compared to the bag-of-genomes approach in the context of natural
260 microbial communities in hot spring mats [Taffs et al., 2009]. A community-wide metabolic model has been
261 reconstructed with KBase and compared with expression data in a bacterial photoautotroph–heterotroph
262 consortium [Henry et al., 2016]. The same method was reused to reconstruct a bag-of-genes model of child
263 gut microbiota for a modelling study of malnutrition-associated dysbiosis [Kumar et al., 2018].

264 Carveme [Machado et al., 2018] provides an alternative between a community-wide GEM and a bag-
265 of-genomes model by assigning a compartment to each species and adding an extracellular compartment
266 to connect them. A community biomass equation is inferred, enabling the use of the constraint-based
267 models described above while keeping the knowledge of metabolite producers. HUMAnN2 can also help
268 navigating between bag-of-genomes and bag-of-genes models by providing both community and species-
269 resolved functional potentials from metagenomic data [Franzosa et al., 2018].

270 **4. Models for large-scale metabolic exploration of microbial** 271 **communities**

272 The recurring goals, when considering a microbial community, are first to identify putative interactions
273 among its members, then to globally assess the role of all species in the community. We focus here on
274 models based on GENREs dedicated to the direct inference of putative interactions, and describe how
275 discrete models of metabolism can be an alternative to constraint-based GEMs for scalable screening of
276 metabolic potential in microbial communities.

277 **4.1 Assessing competition and cooperation potentials of microbial communities**

278 Cooperation through metabolic exchanges and competition for shared nutrients are the main metabolic
279 interactions that occur in microbial communities. As a result, being able to estimate cooperation and
280 competition potentials is essential for characterising, comparing and selecting communities. Approaches
281 dedicated to this purpose can be separated in two categories, the first one composed of methods that
282 structurally analyse and compare individual GENREs, and the second comparing the individual growth
283 rates of bacterial species to those obtained in community using GEMs (Fig. 2 upper right panel).

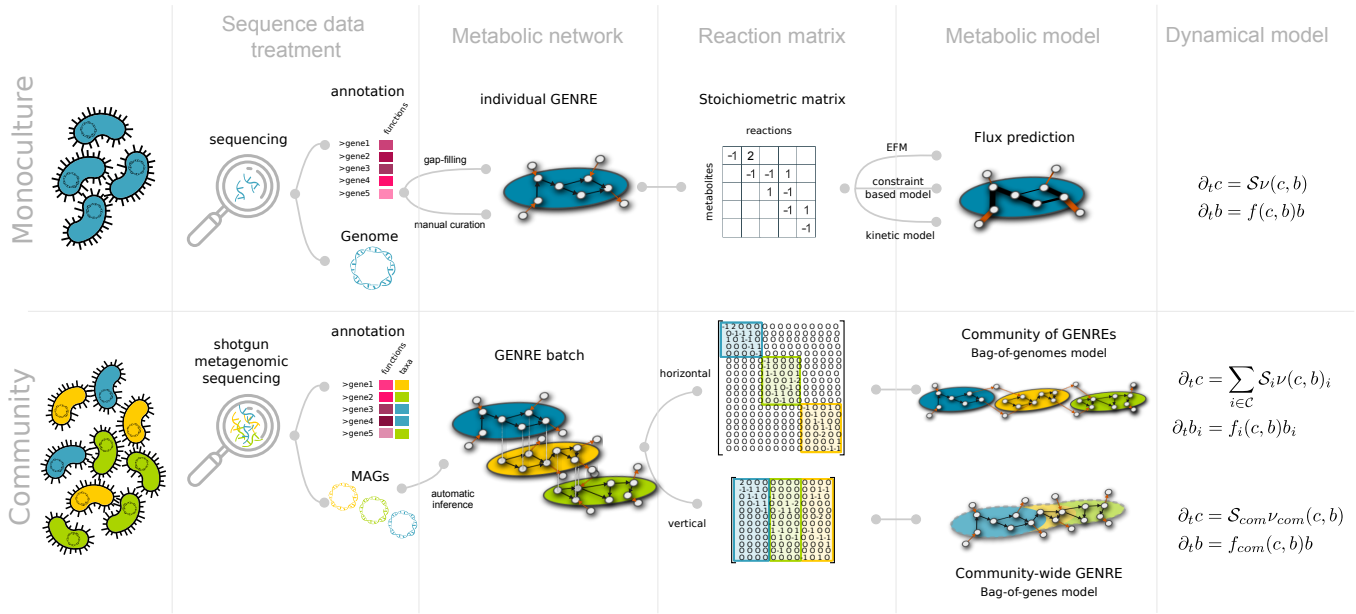


Figure 1. From mono-organism to community function-based model of microbial populations. *Monoculture model* (upper panel): In monoculture, a GENRE is obtained from the annotated genome, with semi-automatic tools (e.g. gap-filling) and manual curation. A matricial representation of the metabolic network, the *reaction matrix*, is an input for flux prediction models. Metabolic fluxes can be computed iteratively to estimate the time evolution of a dynamical model of the population. *Community model* (lower panel): In a microbial community, the metagenome makes it possible to obtain MAGs and GENREs for members of the community. GENREs may share some common functions and metabolites (vertical gray vertices in the GENRE batch). A first paradigm for building a community-wide metabolic model is a *bag-of-genome* model: a set of independent GENREs interacting through common external metabolites. The corresponding reaction matrix is the concatenation of the individual matrices, and the corresponding dynamical model is the sum of the fluxes computed for each unique genome. A second paradigm, termed *bag-of-gene* model consists in merging all the GENREs through their common functions and metabolites, resulting in a global metabolic network for the whole community. The corresponding reaction matrix is the concatenation of the individual matrices after merging or removal of common reactions (columns) and metabolites (rows), resulting in a lower dimension matrix and a dynamical system with a unique equation. With this paradigm analysis can be performed at the scale of a community, but the contribution of each taxon is fuzzier.

284 *Reverse ecology* [Levy and Borenstein, 2012] aims at inferring ecological mechanisms through the char-
285 acterisation of the composition of the environment and the interactions occurring in it. A core step is the
286 detection of exogenously-acquired compounds, or *seeds*, that NetSeed [Borenstein et al., 2008] computes
287 using GENRE’s strongly-connected components. NetCooperate [Levy et al., 2015] quantifies the potential
288 for pairwise syntrophic relationships. The *biosynthetic support score* quantifies the proportion of an en-
289 dosymbiont’s seeds occurring in a host’s GENRE. The *metabolic complementarity index* of two bacterial
290 species is the fraction of one GENRE’s seeds that are products of the second GENRE, and not in its
291 own seed set. Comparing GENREs pairwise can also be used to determine a competition potential. In
292 [Levy and Borenstein, 2013], the *metabolic competition index* is the fraction of an organism’s seeds that
293 also belong to the seeds of another. NetCompt [Kreimer et al., 2012] calculates the *effective metabolic*
294 *overlap*, combining seed calculation with the network expansion algorithm (see section 4.2), and estimating
295 the reduction of a species’s metabolic potential in the absence of source compounds that are shared with
296 other species. These metrics have been used for determining a large-scale evolutionary trend in the loss of
297 biosynthetic capacity over time in parasites [Borenstein and Feldman, 2009], and detecting enzymes oper-
298 ating on exogenously-acquired compounds in diseased individuals [Greenblum et al., 2012]. They were also
299 integrated in analysis frameworks such as RevEcoR [Cao et al., 2016] or PhyloMint [Lam et al., 2020]. It
300 is worth noticing that the seed set is not only the set of nutrients required for growth but includes com-
301 pounds that may theoretically enable the production of any metabolite in the network, and can therefore
302 underestimate or overestimate interaction potentials [Lam et al., 2020].

303 Other methods use the computation of flux distributions with constraint-based modelling to determine
304 scores of competition and cooperation. [Freilich et al., 2011] calculate the *potential competition score* and
305 the *potential cooperation score* to characterise metabolic interactions between nearly 7000 combinations
306 of 118 bacterial GEMs. The authors additionally infer win-lose and give-and-take relationships between
307 bacterial pairs by assessing their growth and co-growth rates in competition- and cooperation-inducing
308 media. MMinte [Mendes-Soares et al., 2016] compares the growth rates of species in pairwise interacting
309 communities to their individual growth rates in order to predict whether one species can benefit from or be
310 negatively impacted by the presence of another. SMETANA overcomes the limitation of pairwise simulation
311 and computes interaction scores for microbial communities by solving mixed integer linear programming
312 problems [Zelezniak et al., 2015]. It computes the *metabolic interaction potential* which represents the
313 extent to which metabolites produced by the community can be reused by species for their growth. The

Predict small-scale community metabolism

Decipher cooperation and competition potentials

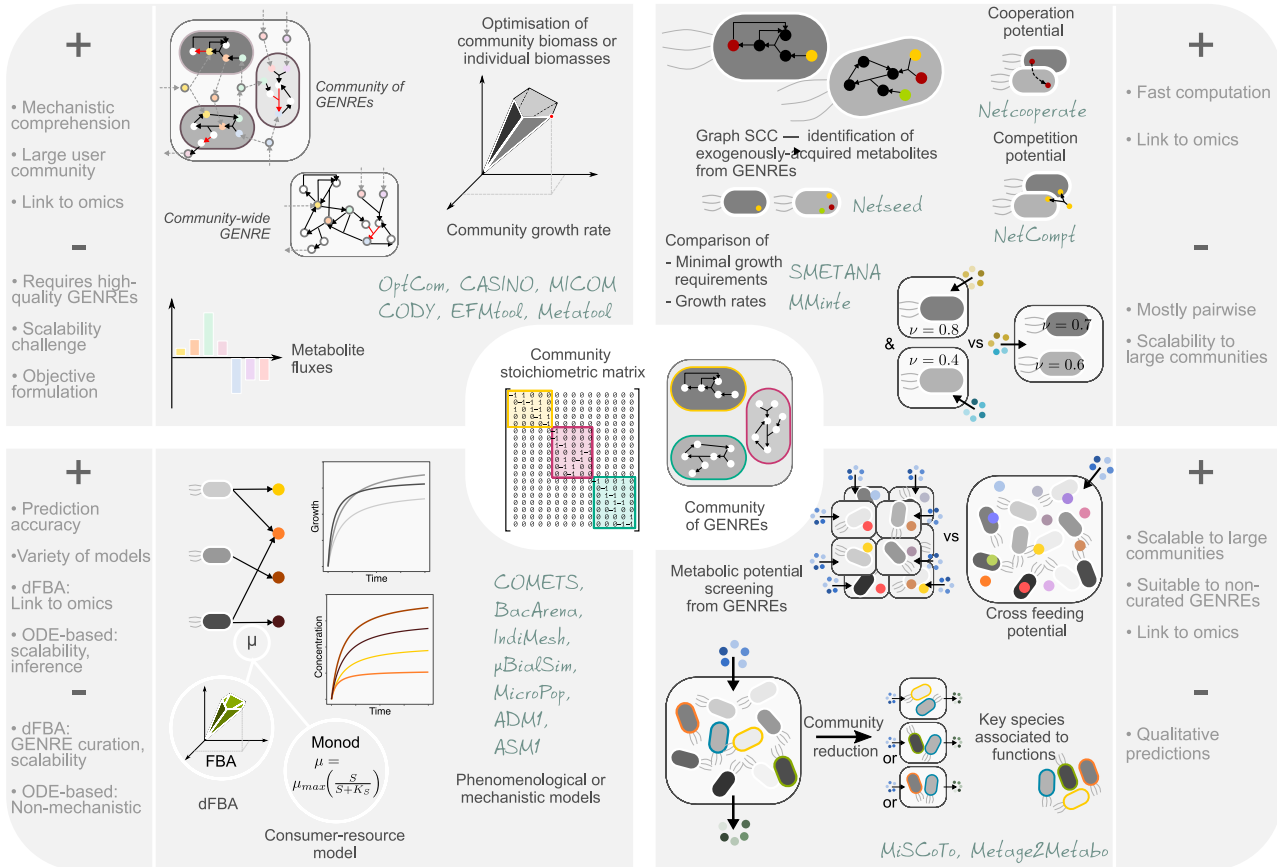


Figure 2. Approaches for numerical simulation and reasoning on community metabolism. Methods and tools are sorted according to the functional characterisation of the microbial community metabolism they can provide: predicting small-scale community metabolism (upper left panel), assessing community dynamics (lower left panel), deciphering cooperation and competition potentials (upper right panel) and characterising large-scale community metabolism (lower right panel). For each class, the main methods and dedicated software (gray italic names) are recapitulated. The main “pros” (+) and “cons” (-) are listed on the lateral box associated to each panel.

314 nutrients required for growth in a non-interacting community can be compared between species, leading
315 to the *metabolic resource overlap*, an estimation of competition potential. [Machado et al., 2021] uses
316 these scores and describes a polarisation of microbial communities between competitive and cooperative
317 metabolisms. Finally, the use of taxon knock-outs in MICOM [Diener et al., 2020] can also be a means to
318 evaluate the positive or negative impacts of species on the growth of other community members.

319 4.2 A Boolean abstraction of metabolic producibility to scale up to large communities

320 As depicted in the above sections, most approaches for modelling interactions in microbial communities
321 are applied pairwise, and more generally, constraint-based models are not applied to large communities
322 due to both the curation needs for the individual GENREs and the computational costs associated with
323 simulations. This motivates the use of alternative paradigms for modelling. One approach is *metabolic*
324 *network percolation* [Bernstein et al., 2019], a probabilistic concept that randomly samples environmental
325 compositions to assess the consistency of metabolite production in microbiomes. Applied to the human
326 oral microbiome, it permitted the creation of an atlas describing metabolite biosynthesis potentials while
327 being more tolerant to missing reactions than the classical constraint-based simulations.

328 Discrete and reasoning-based models are further approaches for estimating the ability of a species
329 to synthesise, or not, metabolic compounds from available substrates in its environments. The *network*
330 *expansion algorithm* [Ebenhöh et al., 2004] provides a Boolean approximation that converts the stoichio-
331 metric matrix into a matrix of incidence of 0, 1 and -1, representing solely the participation of molecules
332 in metabolic reactions, hence discarding mass-conservation. The metabolic potential of a species, called
333 *scope*, is calculated from a set of available nutrients, the *seeds*: metabolic reaction activation depends on the
334 availability of its reactants either from the seeds or as the products of other activated reactions. MeneTools
335 implements network expansion to query metabolic capabilities of GENREs [Aite et al., 2018].

336 Metabolic scope can be used to determine the metabolic potential of several organisms in a community
337 [Christian et al., 2007]. [Ofaim et al., 2017] built a framework considering the relationship between taxon-
338 omy, function and environment, and attempted to use metabolic modelling to explain the co-occurrence
339 of species in the plant rhizosphere or the soil. A similar approach applies to deciphering symbiont-host
340 interactions: [Opatovsky et al., 2018] simulated the metabolism of obligatory and facultative symbionts
341 associated with a phloem-feeding whitefly. As a microbiota can involve hundreds of species living in
342 a shared environment, the scalability of network expansion is a property accounting for the associated

343 metabolic diversity (Fig. 2 bottom right panel). The community-wide implementation of network expansion in MiSCoTo [Frioux et al., 2018] permits computing metabolic potentials of communities and selecting minimal size communities exhibiting functions of interest. Metage2Metabo integrates such analyses together with GENRE reconstruction to systematically compare the metabolic potential of metagenomic samples or genome collections [Belcour et al., 2020]. Discrete models of this kind were used to screen microbial metabolic potentials [Ramalho et al., 2022], to predict potential symbiotic relationships between MAGs from patients with Crohn’s disease and healthy individuals [Sabater et al., 2022], and to analyse metabolic complementarity [Karimi et al., 2021].

351 **5. Towards genomics-informed phenomenological models of microbial** 352 **communities**

353 The use of mechanistic GEMs in the context of large microbial ecosystems can be impeded by poorly-reconstructed GENREs [Babaei et al., 2018], or by issues related to the high dimension of the system, such as uncertainties, accumulations, analysis of massive outputs or computational load. On the other hand, a large number of phenomenological microbial community models, has proved to be efficient in capturing macroscopic features without involving genome-scale inputs. In this section, we discuss opportunities for coupling GEMs with phenomenological and hybrid models to achieve scalability for complex microbiome, while allowing genome-resolved –omics integration.

360 **5.1 Existing models that could benefit from metabolic modelling**

361 Microbial community models other than GEMs have recently been divided into 4 types: **(1)** consumer-resource, **(2)** generalised Lotka-Volterra (gLV), **(3)** trait based models and **(4)** individual based model (IBM) [van den Berg et al., 2022]. Models **(1)** to **(3)** are population-dynamic models based on a system of ordinary differential equations (ODE), describing the rate of change of the population density depending on other species (**(2)** directly, **(1)** indirectly), available resources **(1)**, and environmental factors **(3)**. Model type **(4)** describes the fate of individual cells, whose juxtaposition models the population. IBMs have been used on a wide range of ecosystems in order to gain knowledge of the underlying ecology and for decision making [DeAngelis and Grimm, 2014]. Models of type **(4)** are not opposed to models **(1)-(3)**, as arguably they can be complementary to one another [Hellweger et al., 2016], and since ODE models

370 can be derived as large-population approximations of IBM models [Darrigade et al., 2022]. Due to their
371 computational efficiency, models (1)-(4) can be spatialised, using e.g. partial differential equations (PDEs)
372 as [Labarthe et al., 2019, Moorthy et al., 2015, Smith et al., 2021] did for gut microbiota.

373 GEMs can be considered not only as a category by themselves, but can also be included in models of
374 the form (1)-(4).

375 **Coupling GEMs with consumer-resource models** It must be stressed that bag-of-genomes models
376 (see Fig. 1) can be interpreted as particular cases of consumer-resource models (Fig. 3, Fig. 2 bottom
377 left panel), where resource consumption, metabolite production and growth rates are modelled through
378 FBA fluxes. Hence, existing consumer-resource models can be easily supplemented with a genome-scale
379 metabolic model. As such, a consumer-resource model that qualitatively reproduces experimentally ob-
380 served ecological patterns of the Earth Microbiome Project by considering the exchange of secondary
381 metabolites [Marsland et al., 2020] could be refined with GEMs: export fluxes for the metabolites included
382 in the consumer-resource model could be computed dynamically with the GEM while remaining essential
383 metabolites would be given a steady-state value.

384 **Deciphering gLV interactions with GEMs** gLV models proved to be efficient in analysing interac-
385 tions across the gut microbiome by fitting time-series of population counts [Stein et al., 2013]. Population
386 data obtained from high-throughput sequencing reports relative abundances, inducing inference difficul-
387 ties [Remien et al., 2021]. That drawback can be overcome by providing *priors* to the inference methods
388 [van den Berg et al., 2022], for instance from GEMs. In an example coming from anaerobic digestion (AD),
389 GEMs were reconstructed for each of the 800 species present in the community, and FBA was used to under-
390 stand the syntrophic relations between micro-organisms [Basile et al., 2020], that could be used as *priors*
391 of gLV model parameters. Conversely, a pairwise interaction identified in time-series with a gLV model
392 could be interpreted using pairwise GEMs.

393 **Supplementing trait-based models with GEMs.** Bioreactor modelling is a good example that re-
394 lates to both trait-based models (since they often involve abiotic parameters such as pH or temperature)
395 and consumer-resource models. Two examples of widely implemented bioreactors are wastewater treat-
396 ment plants based on activated sludge (AS) [Tchobanoglous et al., 2003], and biogas production from AD
397 of organic waste [Moletta, 2015]. AS [Leslie Grady Jr et al., 1986] and AD [Batstone et al., 2002] systems

398 have been modelled in the framework of community-wide ODE systems (see Fig. 1), representing micro-
399 bial species as functional guilds, i.e. a population of microbes responsible for a metabolic pathway. AD
400 models have been used in various contexts such as ruminal [Muñoz-Tamayo et al., 2016], environmental
401 [Desmond-Le Quémener et al., 2021] or gut microbiota [Kettle et al., 2018, Muñoz-Tamayo et al., 2010].
402 Since considerable efforts have been made to sequence microbes present in AD and AS at a global scale
403 [Dueholm et al., 2022], one could readily envision enhancing parts of the AD and AS with metabolic mod-
404 els. However, FBA analysis alone may not suffice to represent abiotic inhibition arising in AD systems
405 [Basile et al., 2023b].

406 **Coupling IBM with GEMs.** In an IBM model, each individual can be provided with a GEM, which can
407 be used to predict its metabolism with FBA. Such a framework has been already included in platforms like
408 BacArena [Bauer et al., 2017], ACBM [Karimian and Motamedian, 2020], IndiMeSH [Borer et al., 2019] or
409 CROMICS [Angeles-Martinez and Hatzimanikatis, 2021], that were recently compared [Scott et al., 2023].
410 These frameworks allow spatialisation: the IBM can include movement functions modelling microbial
411 motility, and complex multi-physics features, e.g. the complex diffusive pattern in the rhizospheric area for
412 IndiMesh.

413 Since coupling system dynamics, IBM and GEMs involve different mathematical formalisms, hybrid
414 models provide a unifying framework that can help implementing model coupling, as described in the next
415 section.

416 5.2 Extending the scope of community modelling with hybrid models

417 The complex behaviours of microbiomes arise from the interaction of different functional processes with
418 different dynamics, that can be considered as continuous, discrete or random, depending on their time
419 scale, dictating appropriate mathematical formalisms [Assar et al., 2014]: *numerical ODEs* for continuous
420 dynamics, *automata* for discrete instantaneous changes, *nondeterministic* or *stochastic models* for uncertain
421 or random transitions. This highlights the need for a flexible framework to explore conjointly these processes
422 at the community scale.

423 *Hybrid models* are such a framework, by assembling modules modelled using different formalisms. In
424 [Liu et al., 2022] five complementary combinations of modelling formalisms are used for biological sys-
425 tems: *hybrid discre-continuous* methods, often used for intracellular modelling as above; *hybrid stochastic-*
426 *deterministic* methods, introducing random delays; *hybrid flux balance analysis* methods, integrating

427 metabolic, regulatory, and signalling networks; *hybrid logic-quantitative* methods, combining quantitative
428 methods with Boolean networks or fuzzy logic; and *spatial hybrid* methods, combining spatial stochastic
429 and deterministic models to bridge three scales: time, abundance, and space [Widmer and Stelling, 2018].

430 More formally, *hybrid dynamical systems* often combine continuous *state* and *control* variables plus
431 a discrete *mode* variable, the transition of which is determined by *guard* conditions defined on the con-
432 tinuous variables [Henzinger, 1996]. In *switched systems*, the mode activates, deactivates, and combines
433 modules of a hierarchical model [Branicky, 1994, Shorten et al., 2007]. In this framework, phenomenolog-
434 ical models and GEMs can be coupled in a flexible and rigorous way. A good example can be found in
435 [Assar et al., 2014], that switches between two ODE models [Coleman et al., 2007, Scaglia et al., 2009] and
436 a dFBA model of wine fermentation kinetics [Pizarro et al., 2007] according to the system state.

437 **6. From environmental samples to functional and taxonomic diversity**

438 The preceding has taken a high-level view of the modelling process and the insights models can provide,
439 and has assumed that the collection of models for a community accurately reflects the taxonomic com-
440 position of its members and the biomolecular functions they provide. These will vary greatly between
441 environments and from sample to sample: different species provide different functions, and within those
442 species strain differences and horizontal gene transfer will further affect gene content. In this section we
443 consider how molecular identification of taxonomic and functional diversity is employed to build accurate
444 characterisations of real communities.

445 Two DNA sequencing approaches are used when building metabolic models from environmental sam-
446 ples [Pérez-Cobas et al., 2020]. *Amplicon sequencing* of marker genes characterises taxonomic diversity
447 within a sample and can be used to infer functional profiles, map community members to existing mod-
448 els [Heinken and Thiele, 2022b], and select genomes prior to GENRE reconstruction. *Whole metagenome*
449 *shotgun sequencing* characterises both taxonomic and functional diversity within a sample by determin-
450 ing the gene content of community members, and can be used to infer new GENREs of novel species
451 and strains. In both approaches, *third generation sequencing* (TGS) using long reads can complement, or
452 advantageously replace, NGS using short reads.

453 Amplicon sequencing and shotgun metagenomic sequencing have demonstrated the high diversity of
454 microbial communities, for bacteria and archaea, but also eukaryotes such as fungi [Tedersoo et al., 2014].

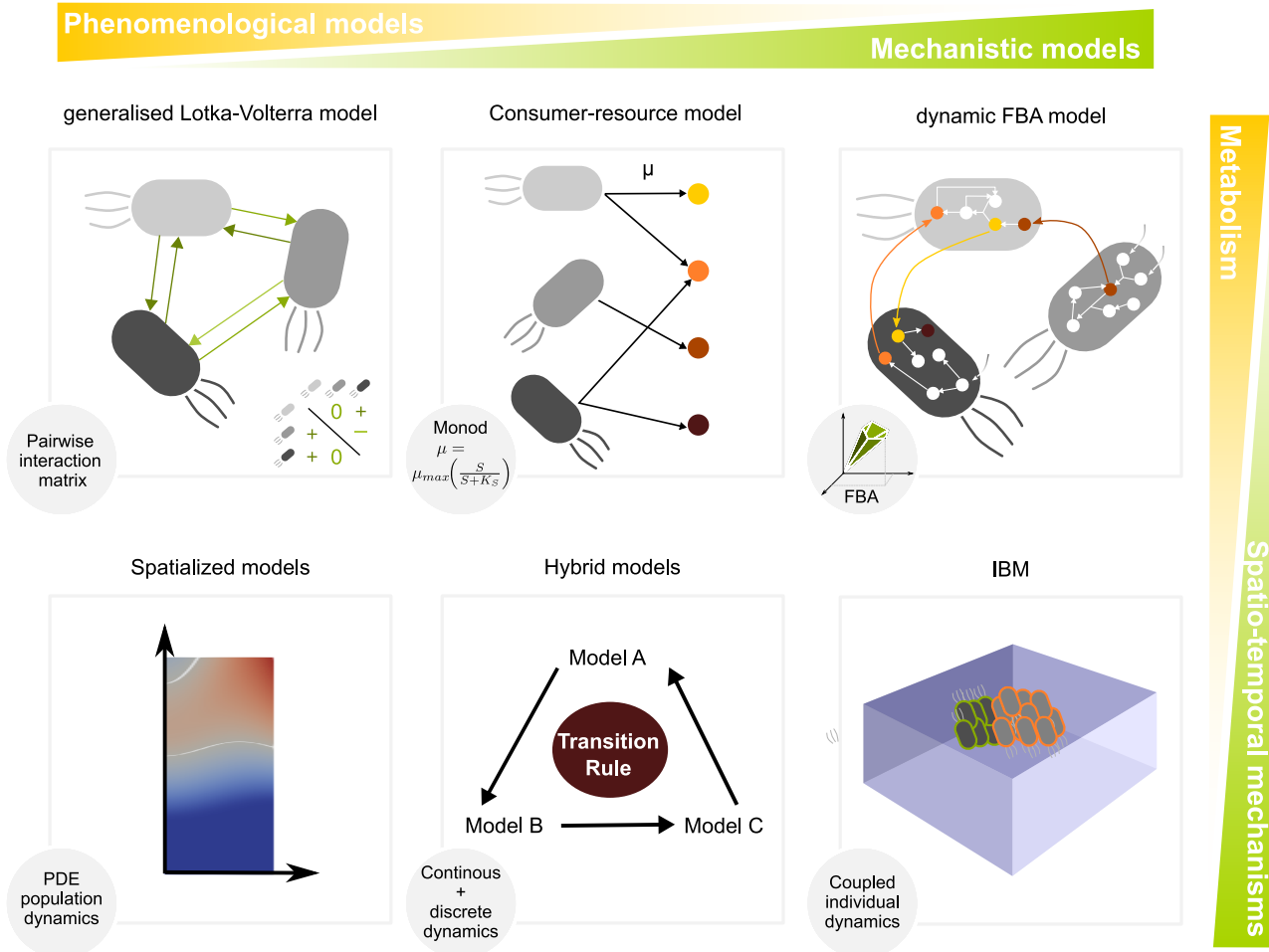


Figure 3. Dynamic models of metabolism: opportunities beyond GENRE-based models. Different formalisms can be used to capture microbial population dynamics (upper panel). These models can be placed on a gradient, from phenomenological (left) to mechanistic models (right) of the metabolic capabilities of the community. Each formalism can be supplemented by additional spatio-temporal mechanisms (lower panel) that can be represented at the population (PDE) or individual scale (IBM). The addition of spatio-temporal features in the community dynamics induces simplifications in the metabolic models to keep the model computationally tractable.

Table 3. Main short read and long read metagenomic sequencing technologies. Abbrev. PacBio, Pacific Biosciences; ONT, Oxford Nanopore Technologies; bp, base pair; Q, quality score logarithmically related to the base calling error probability, Q20: 1%, Q30: 1‰ etc. Read length is subject to variation.

Method	sequencing platform	read lengths (bp)	throughput	specifications	applications advantages/pitfalls
short-read metagenomics	Illumina Miseq system	2x300	low	1-25M reads, up to 15 Gbp >70% bases higher than Q30	- Taxonomic profiling, - Relatively low data processing requirements
	Illumina Novaseq 6000	2x150	high	1.6B - 40B reads, up to 3 Tb, ≥ 85% of bases higher than Q30	- Taxonomic and functional profiling, - Genome assembly and binning; - High throughput, low cost for production-scale genomics
	Pacific Biosciences Onso	2x150	moderate	120-150 Gbp ≥90% bases Q40	- Taxonomic and functional profiling - Genome assembly and binning - Identification of variants - 15x higher accuracy than other short read platforms - High accuracy in characterisation of homopolymers - Higher costs/bp
long-read metagenomics	Pacific Biosciences Sequel II and IIE	typically 10k-30k	moderate	3M reads typically Q20-Q40 reads	- Taxonomic and functional profiling, - Genome assembly and binning - Can reconstruct circular MAGs - High sample input: ~2 µg - Full-length 16S rRNA seq
	Pacific Biosciences Revio system	~10k to ~30k	moderate	up to 90 Gbp Q30-Q40	- Higher HiFi read accuracy with DeepConsensus base calling, methylation calling - High sample input: ~2 µg - Full-length 16S rRNA seq
	Oxford Nanopore Technologies Promethion R10 series	on average ~10k to ~60k, strongly dependent on DNA extraction	moderate	~50 Gb > median Q20 most reads, ~Q30 (99.9%) for Duplex reads	- Very long reads - Transportable technology - Low sample input: ~1 µg - Less polishing needed than R9 series - Read quality still lower than PacBio

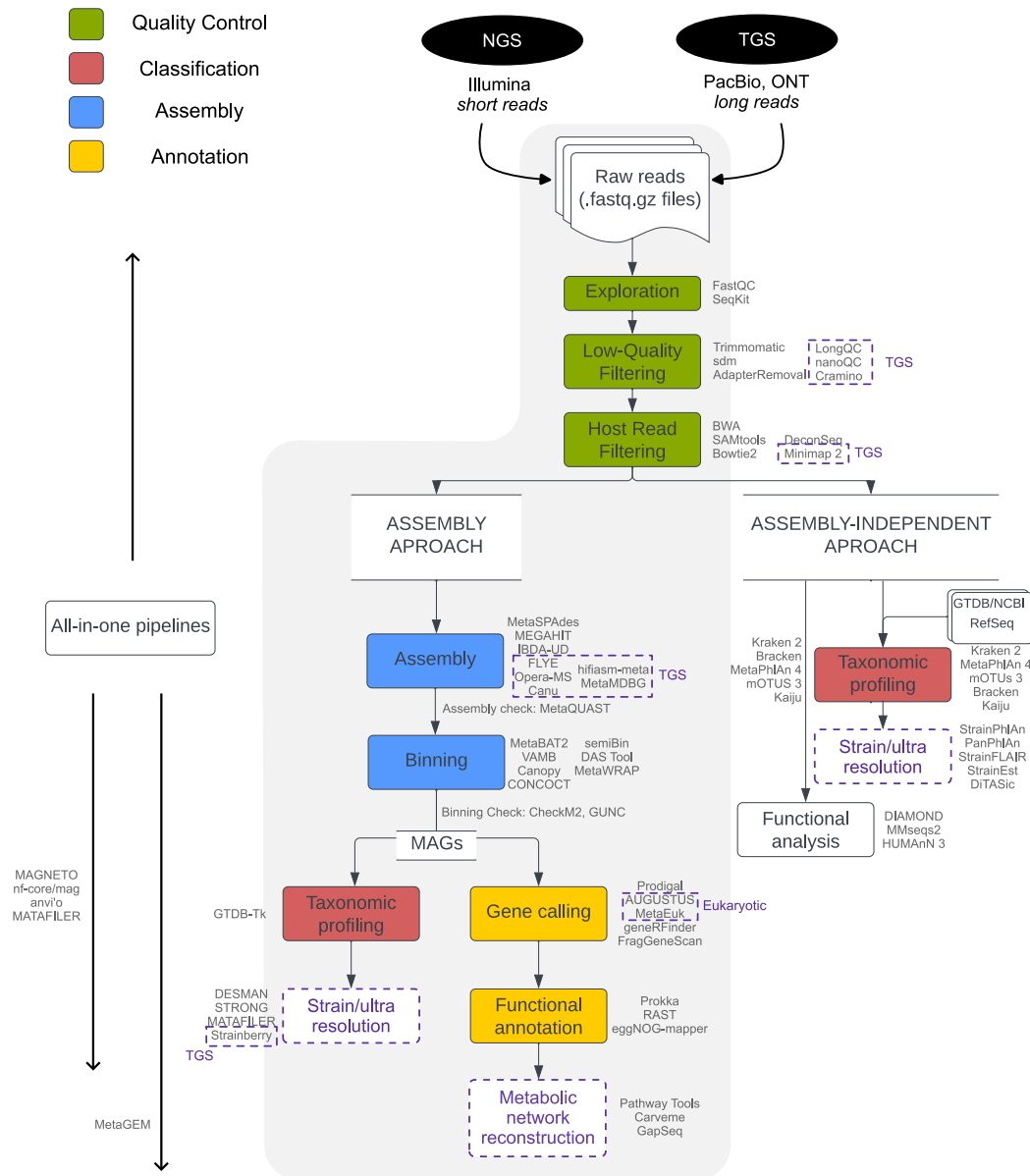


Figure 4. Metagenomic sequencing and data processing. After quality control, shotgun metagenomic reads can enter an assembly pipeline or an assembly-independent profiling pipeline. The assembly pipeline is highlighted as it permits the reference-free reconstruction of MAGs and subsequently GENRE reconstruction that directly survey the functions present in the ecosystem. Examples of tools associated to each step are indicated next to the box. Tools or steps related specifically to TGS, eukaryotic metagenomics and strain resolution are highlighted with dotted boxes.

455 **6.1 Amplicon sequencing**

456 A good *marker gene* is universally present in target populations and taxonomically robust because its in-
457 frequent mutation is not biased by functional selection. Typical choices for marker genes are 16S ribosomal
458 RNA genes in prokaryotes, and either 18S ribosomal RNA genes or different lineage-specific markers in eu-
459 karyotes [Earl et al., 2018, Johnson et al., 2019, Bradley et al., 2016, Burki et al., 2020, Tekle et al., 2010].
460 For a given environmental sample, the variable regions of the selected marker genes are amplified using
461 PCR and later sequenced. Until recently, inexpensive short-read NGS sequencing has been used exclusively
462 [Kim et al., 2021, Bukin et al., 2019], although this limits the size of the variable region that can be used,
463 may be subject to amplification bias, and may limit taxonomic resolution to phylum and genus due to 16S
464 polymorphism [Knight et al., 2018, Mitchell et al., 2020, Sieber et al., 2018a]. Long-read TGS sequencing
465 of full-length 16S rRNA has recently been shown to reduce bias, broaden the choice of marker genes,
466 and increase taxonomic resolution to species and subspecies [de Muinck et al., 2017, Johnson et al., 2019,
467 de Oliveira Martins et al., 2019].

468 Amplicon sequences are clustered into *operational taxonomic units* (OTUs) or into *amplicon sequencing*
469 *variants* (ASVs) that provide “species-like” delineation of organisms and estimations of relative abundance.
470 Bioinformatics pipelines for amplicon sequencing analysis are reviewed in [Hakimzadeh et al., 2023]. While
471 the functions present in a microbiome cannot be directly identified by amplicon sequencing [Almeida et al., 2021],
472 approximations may be indirectly inferred using phylogenetic placement of markers [Douglas et al., 2020,
473 Bowman and Ducklow, 2015], alignment [Wemheuer et al., 2020, Narayan et al., 2020], or incorporation of
474 biological knowledge in the form of taxonomic and functional profiles [Mallick et al., 2019, Jing et al., 2021].
475 Nonetheless, limitations resulting from the importance of the accessory genome in a species with respect
476 to its core genome will impact the accuracy of the predictions.

477 Building mechanistic models of metabolism requires accurate taxonomic classification, both for identi-
478 fying existing genomes and GENREs and characterising diversity. The KSGP database [Grant et al., 2023]
479 combines the GTDB unified taxonomy [Parks et al., 2020] with a collection of environmentally sampled
480 full-length sequences, to substantially increase taxonomic assignment of environmental samples and links
481 to reference genomes. Resolution within reference genomes is typically limited to species and will miss
482 most strain-specific variability.

Table 4. Main sequence analysis methods.

OTU clustering §6.1		Long read assembly §7.2	
QIIME 2	[Estaki et al., 2020]	MetaFlye	[Kolmogorov et al., 2020]
DADA2	[Callahan et al., 2016]	Canu	[Koren et al., 2017]
LotuS2	[Özkurt et al., 2022]	Opera-MS	[Bertrand et al., 2019]
Function inference §6.1		hifiasm-meta	[Feng et al., 2022]
PICRUSt2	[Douglas et al., 2020]	MetaMDBG	[Benoit et al., 2023]
paprica	[Bowman and Ducklow, 2015]	Assembly evaluation §7.2	
Tax4Fun2	[Wemheuer et al., 2020]	MetaQUAST	[Mikheenko et al., 2016]
Piphillin	[Narayan et al., 2020]	Gene prediction §7.2.1	
Metabolic features §6.1		Prodigal	[Hyatt et al., 2010]
MelonnPan	[Mallick et al., 2019]	MetaEuk	[Karin et al., 2020]
Meta-Apo	[Jing et al., 2021]	Reference gene catalogues §7.2.1	
Taxonomic classification §6.1		CD-HIT	[Fu et al., 2012]
KSGP	[Grant et al., 2023]	MMseqs2	[Steinegger and Söding, 2017]
GTDB	[Parks et al., 2020]	Binning §7.2.2	
Adapter removal §7.1		CONCOCT	[Alneberg et al., 2014]
Trimmomatic	[Bolger et al., 2014]	MaxBin 2	[Wu et al., 2016]
sdm	[Hildebrand et al., 2014]	MetaBAT 2	[Kang et al., 2019]
AdapterRemoval	[Schubert et al., 2016]	Bin consolidation §7.2.2	
Quality control §7.1		VAMB	[Nissen et al., 2021]
FastQC	[Andrews et al., 2010]	SemiBin2	[Pan et al., 2022]
SeqKit	[Shen et al., 2016]	MetaWRAP	[Uritskiy et al., 2018]
Long read analysis §7.1		DAS Tool	[Sieber et al., 2018b]
LongQC	[Fukasawa et al., 2020]	Chimerism and contamination §7.2.2	
nanoQC	[Coster et al., 2018]	CheckM	[Parks et al., 2015]
Cramino	[Coster and Rademakers, 2023]	GUNC	[Orakov et al., 2021]
Minimap2	[Li, 2018]	MAG creation pipelines §7.2.2	
Taxonomic filtering §7.1		MAGNETO	[Churchward et al., 2022]
MetaPhlan4	[Blanco-Miguez et al., 2022]	nf-core/mag	[Krakau et al., 2022]
mOTUs3	[Ruscheweyh et al., 2022]	anvi'o	[Eren et al., 2021]
Functional profiling §7.1		MATAFILER	[Hildebrand et al., 2019]
Kaiju	[Menzel et al., 2016]	MetaGEM	[Zorrilla et al., 2021]
DIAMOND	[Buchfink et al., 2014]	Short read assembly §7.2	
MMseqs2	[Steinegger and Söding, 2017]	metaSPAdes	[Nurk et al., 2017]
Short read assembly §7.2		IDBA-UD	[Peng et al., 2012]
metaSPAdes	[Nurk et al., 2017]	MEGAHIT	[Li et al., 2015]
IDBA-UD	[Peng et al., 2012]		
MEGAHIT	[Li et al., 2015]		

483 6.2 Shotgun metagenomics

484 Whole metagenomic sequencing uses total DNA extracted from the environmental sample to build a com-
485 plete picture of the genomes of the organisms present in the sample. Also called *untargeted*, *shotgun*,
486 *random*, or *WGS* sequencing, whole metagenomic sequencing does not target a specific amplified marker
487 gene, and sequencing a larger amount of genetic material comes with increased expense. However, it avoids
488 many of the shortcomings of amplicon sequencing: it can taxonomically resolve species, and even strains
489 with sufficient data; it can identify the genes carried by species on a sample-specific basis, even for un-
490 known taxa; and it can infer microbial functions including strain-specific variants in the accessory genome
491 [Djemiel et al., 2022, Hildebrand, 2021, Hildebrand et al., 2019, Hildebrand et al., 2021].

492 Table 3 summarises the short-read and long-read sequencing technologies used for assembling genomes,
493 and reconstructing metagenomes and MAGs [MetaHIT consortium et al., 2015, Nielsen et al., 2014, Pan et al., 2023].
494 As above, long TGS reads can significantly improve results [Kim et al., 2022]. PacBio HiFi long read se-
495 quencing has low error rates and provides near-finished microbial genomes from pure cultures or metagenomes,
496 albeit at relatively high cost [Amarasinghe et al., 2020]. Oxford Nanopore (ONT) sequencing produces even
497 longer reads at higher throughput on average, but with higher rates of homopolymer indels that may lead
498 to gene calling errors [Watson and Warr, 2019]. Combining long-read ONT sequencing with short reads for
499 post-assembly error correction is a common compromise [Delahaye and Nicolas, 2021], and more recently
500 PacBio metagenomic assembly using metaMDBG [Benoit et al., 2023] has improved on that. Both PacBio
501 and ONT are rapidly advancing their technological development, and due to improvements in sequence
502 length, quality, and throughput will likely replace short read NGS technologies entirely in the near term.

503 6.3 Metagenomic strains

504 In order to fully understand microbial communities, it is necessary to resolve them into individual strains,
505 the fundamental unit of microbiological diversity [Segata, 2018, Hildebrand, 2021]. *Escherichia coli*, which
506 has a highly variable genome, is a well-known example of the importance of strains in functional char-
507 acterisation: some are harmless commensals while others are harmful pathogens [Leimbach et al., 2013].
508 Resolving metagenomic data at the strain level brings a new dimension to functional prediction and to the
509 comparison of the metabolic potential of samples. A *metagenomic strain* is a set of clonal descendants of a
510 single cell, with such low levels of recombination with other strains that it can be distinguished genetically
511 from them. Members of a metagenome strain may differ substantially, particularly in the rapidly evolving

512 accessory genome [Maistrenko et al., 2020]. Reconstructing metagenomic strains is difficult since sequenc-
513 ing depth may be insufficient for genome reconstruction and conspecific strains with overlapping genetic
514 regions may be present. Strain-resolved genomes can be determined by sequencing cultured isolates or
515 single cells, but the former are not representative of the community and the latter remain too expensive
516 and low-throughput for many applications. For these reasons, there is a practical need for efficient methods
517 that can profile microbial communities at high genomic resolution [Bickhart et al., 2022].

518 *Reference-based methods* for delineating microbial strains in metagenomes rely in general on aligning
519 metagenomic reads to a reference database and inferring strain identity based on single nucleotide polymor-
520 phisms (SNPs) [Scholz et al., 2016, Truong et al., 2017, Albanese and Donati, 2017, Fischer et al., 2017,
521 Silva et al., 2021]. These methods identify gene family presence, strain identification for dominant strains,
522 and abundance estimates.

523 *Strain-resolved metagenomics* compares *de novo* reconstructed MAGs from different metagenomes to
524 each other, without use of a reference database. Gene catalogues can be used to trace core genes in MAGs
525 across samples, which can be compared to create within-species strain-resolved phylogenies of metagenomes
526 and avoid pan-genome-induced phylogenetic idiosyncrasies [Hildebrand et al., 2021].

527 When investigating conspecific strains in the same metagenomic sample, specialised tools model SNP
528 frequencies across assemblies to infer bacterial haplotypes and identify conspecific strains, even to provid-
529 ing nearly complete pan genome reconstructions, or delineate strains using TGS data of low-complexity
530 metagenomes [Quince et al., 2017a, Quince et al., 2021, Vicedomini et al., 2021]. Metagenomic strains can
531 be mapped to GENREs, as shown in [Basile et al., 2023a] where longitudinal community models were built
532 to describe metabolic changes over time of a well sampled microbiome.

533 Single-cell technologies can also rise to the challenge of recovering genomes of low abundant strains
534 and microbial dark matter (see [Lloréns-Rico et al., 2022] for review), especially through the genera-
535 tion of *single amplified genomes* (SAGs) and the promise of modelling rare species [Bowers et al., 2022,
536 Lyalina et al., 2022, Zheng et al., 2022, Arikawa et al., 2021]. Single-cell transcriptomics also promises to
537 contribute to context-specific models [Hrovatin et al., 2021].

7. From metagenomic data to assembled genomes

7.1 First steps of data processing and assembly-free analysis

A standard metagenomic analysis workflow comprises: (1) sample collection, DNA extraction, and DNA sequencing; (2) data pre-processing and read curation, including quality control and filtering; (3) taxonomic profiling via either “read-based” or “assembly-based” approaches; and (4) further downstream analyses of genomic and functional characteristics, such as gene prediction, and functional or taxonomic annotation [Quince et al., 2017b] (Fig. 4). In this simplified workflow, (2), (3) and (4) are typically executed in a single pipeline, or can be adjusted to the analytical needs in a project. A crucial decision in the bioinformatics analysis is the choice between reliance on assemblies, or on analysing metagenomes independent of this. Assemblies will in most cases require further computational resources and user expertise, but offer in exchange a much richer datascape to explore in the follow up analysis. For example, to build models of metabolism and GENREs in particular, sequence assembly and MAG reconstruction are required to delineate the functional role of bacterial species in the community.

Once sequences for the chosen samples have been acquired, they are screened; trimmed to remove low-quality ends, bar codes, adaptors and technical artefacts; and filtered for quality (Table 4). Specific tools are available for quality filtering of long-read sequencing data [Fukasawa et al., 2020, Coster et al., 2018, Coster and Rademakers, 2023, Li, 2018]. Host-associated samples such as gut and oral metagenomes may be contaminated with reads originating in the host organism, that should be removed to avoid chimeric assemblies or false-positive assignment to bacterial taxa [Li et al., 2009, Li and Durbin, 2009, Schmieder and Edwards, 2011].

While assembly-based metagenomics is the approach of choice for building mechanistic models of metabolism, we briefly present here assembly-independent alternatives for taxonomic functional inference. Pre-processed metagenomic reads can also be used to infer the microbial taxonomic and functional potential composition in the sequenced samples, using DNA, marker gene, or protein-based approaches. Read-based taxonomic profilers estimate the relative abundance of microbes in an environmental sample by aligning the pre-processed sequencing reads to a reference sequence database such as GTDB [Parks et al., 2020] or NCBI RefSeq [O’Leary et al., 2016], using either mapping [Langmead and Salzberg, 2012] or k-mer matching [Lu et al., 2022]. The reference genome search space can be drastically reduced by including only conserved marker genes that are specific to certain taxonomic groups [Blanco-Miguez et al., 2022, Ruscheweyh et al., 2022]. Read-based profilers can also be used for functional inference [Beghini et al., 2021].

567 Protein-based functional profilers use 6-frame translations into amino acid sequences [Menzel et al., 2016,
568 Buchfink et al., 2014, Steinegger and Söding, 2017], with the advantage of greater sensitivity because of
569 the lower mutation rates in amino acids, but at the cost of significant increase in run time [Ye et al., 2019].

570 The choice of reference database plays an important role in read-based taxonomic profiling. Microbes
571 lacking a reference genome will remain undetected, which is a problem in metagenomics as much of the
572 microbial diversity is not yet represented in reference databases [Solden et al., 2016]. In addition, the read-
573 based approach does not permit the direct association of microbial diversity to metabolic functions. While
574 it can be used for building a community-wide model as a super-organism associated to a metagenome, a
575 common protocol is instead to assemble the reads and reconstruct genomes.

576 7.2 Metagenomic assemblies

577 Reconstructing microbial genes or partial genomes by assembling sequences from environmental samples
578 overcomes limitations of reference-based metagenomic approaches, but requires significant sequencing read
579 depth and bioinformatics expertise. While shallow shotgun sequencing (*ca.* 0.5 million short reads per
580 sample) can be enough to capture the taxonomic and functional diversity of microbiome samples at rel-
581 atively low cost, it does not provide sufficient coverage for *de novo* assembly of novel genes and genomes
582 [Lapidus and Korobeynikov, 2021].

583 Quality filtered reads can be directly assembled using metagenomic-specific assemblers, designed for
584 short reads, long TGS reads, or in hybrid combinations (Table 4), see [Brown et al., 2021] for review.
585 Assemblers that specifically target high-fidelity long reads perform well in general on middling-complex
586 communities such as the gut microbiome, and have an acceptable computational efficiency for large scale
587 metagenomes [Feng et al., 2022, Benoit et al., 2023]. When sampled metagenomes are expected to be
588 highly similar to each other, as in *e.g.* time series data of the same ecosystem with reasonably stable
589 strain persistence, co-assemblies of time series samples can be performed to increase assembly quality
590 [Hildebrand et al., 2019, Hildebrand et al., 2021].

591 The drawback of assembly-based metagenomics is the high potential for errors in the assembly, and
592 the inability to include a large part of the reads in the assembly, especially those from low-abundant
593 genomes in highly diverse, complex ecosystems. The fraction of non-assembled reads depends on several
594 factors, sequencing depth and microbiome diversity especially, and can range from <5% in gut, to >90%
595 in soil metagenomes [Hildebrand et al., 2019, Bahram et al., 2018]. Even with deeper short-read shotgun

596 sequencing, *de novo* assembly remains a challenge due to the presence of repetitive DNA regions, shared
597 genomic regions between strains, and the complexity of microbial communities with unknown and uneven
598 representation of both diverse and closely-related strains [Somerville et al., 2019, Ayling et al., 2020]. TGS
599 can overcome many of the difficulties in metagenomic assembly by spanning repetitive sequences and highly-
600 conserved regions. Before the development of the latest low-error technologies, the need for short-read data
601 to overcome high error rates and low coverage made this approach costly and complex [Kim et al., 2022].
602 The low error rate of PacBio HiFi reads constitutes a valuable alternative with high quality, circular
603 genomes frequently being reported in the data currently available (e.g. [Kim et al., 2022]), (Table 3).

604 7.2.1 Gene catalogue creation and gene prediction

605 Genes can be predicted on contigs, but having the complete sequence of a gene and its context makes its
606 functional and taxonomic assignment and its annotation much easier and more reliable, facilitating the
607 construction of community-wide GENREs [Tamames and Puente-Sánchez, 2018].

608 To capture the set of proteins represented in a set of metagenomes, one can generate reference gene
609 catalogues. These typically start with the genes predicted on contigs from metagenomic assemblies, that
610 are subsequently clustered e.g. at 95% identity, in order to remove trivial differences between sequences
611 due to fragmentary data (e.g. genes that miss the start or stop codons), sequencing errors or small, strain-
612 level variations [Qin et al., 2010]. The clustering can be performed either at the nucleotide or amino acid
613 level [Fu et al., 2012, Steinegger and Söding, 2017]; the former provides greater resolution for taxonomic
614 classification, whereas the latter is better suited for identifying functionally-relevant similarities between
615 distant orthologues [Commichaux et al., 2021, Wang and Jia, 2016].

616 The underlying goal of the clustering process is to reproducibly group together sequences that have the
617 same function or the same taxonomic origin, thereby identifying the genes from which the sequences are
618 derived in a way that is consistent across samples. Each cluster is typically represented by one sequence,
619 either a representative selected from the sequences clustered together, or a sequence that represents the
620 consensus of the clustered sequences. Cluster representatives can be used in a broad range of sequence-
621 based analyses (*e.g.* database searches, function or structure predictions), and also to estimate the rel-
622 ative abundance of a single gene across metagenomes using a gene catalogue [Commichaux et al., 2021,
623 Plaza Oñate et al., 2019]. As gene catalogues reduce the redundancy of the data, they can be used as a
624 basis of exhaustive surveys that compare the metabolic potential of metagenomic samples by associating

625 metabolic functions to the genes.

626 7.2.2 Creation of metagenome-assembled genomes (MAGs)

627 Most contigs obtained from assemblies are only fragments of pro- or eukaryotic genomes. TGS permits the
628 assembly of near-complete genomes for abundant species. For less abundant organisms and when using
629 short read sequencing data, it is necessary to add a *binning step* that sort contigs into “bins” according to
630 the genome they likely originate from, based on assumptions of shared sequence composition, orthologous
631 linkage data, and shared co-abundance within and between samples [Alneberg et al., 2014, Wu et al., 2016,
632 Kang et al., 2019, Nissen et al., 2021, Pan et al., 2022]. VAMB [Nissen et al., 2021] in particular allows
633 splitting clusters into sample-specific bins according to their sample of origin, an approach termed “multi-
634 split,” which is useful for comparing pan genomes of conspecific strains. *Bin consolidation* assimilates bins
635 created by different tools into a set of refined bins [Uritskiy et al., 2018, Sieber et al., 2018b].

636 Genome bins of sufficient quality to approximate a genome are commonly called “MAGs,” for *metagenome-*
637 *assembled genome*. Popular measures of MAG quality are the genome *completeness*—a desirable measure
638 of the proportion of genome recovered for the species of interest—and genome *contamination*, the undesir-
639 able amount of genome fragments that are of different species origin and been mistakenly binned together
640 [Chen et al., 2020]. Tools exist for quality assessment using single-copy marker genes as well as detection
641 of contamination and chimeric assemblies (Table 4). A MAG is said to be of high quality when it has
642 over 90% of the expected count of SCGs with less than 5% redundancy of their prevalence, and containing
643 assemblies of ribosomal RNA (rRNA) and transfer RNA (tRNA) genes [Bowers et al., 2017]. Abundance
644 of MAGs in a sample can be estimated using (1) mapping raw metagenomic reads on the newly created
645 MAG and (2) extracting informative marker genes for each MAG to which then either reads are mapped,
646 or their abundance is derived from a gene catalogue (as used in [Hildebrand et al., 2021]).

647 Even though building MAGs is data- and computation-intensive, they provide significant insights into
648 the genomic and functional diversity found in different habitats [Almeida et al., 2019, Kashaf et al., 2022].
649 MAGs are contig-based clusters and thus preserve the relative order of genes along contigs, which is of
650 importance as genes found in close physical proximity may encode biosynthetic pathways for the synthesis
651 of specialised metabolites [Almeida et al., 2021]. Dozens of high quality MAGs can be reconstructed from
652 short-read data of a single gut metagenome, but rarely as continuous, circular genomes. Here, TGS tech-
653 nologies can potentially yield dozens or hundreds of closed genomes without binning, markedly increasing

654 the quality of culture-independent genome reconstructions [Kim et al., 2022].

655 All-in-one pipelines combine tools to reconstruct MAGs directly from sequencing reads [Churchward et al., 2022,
656 Krakau et al., 2022, Eren et al., 2021, Hildebrand et al., 2019]. MAGs can be used for GENRE recon-
657 struction and thus enable metabolic modelling, as in [Bernardini et al., 2022] where the authors per-
658 formed FBA simulations to examine interspecies metabolic interactions in anaerobic digestion. MetaGEM
659 [Zorrilla et al., 2021] is a MAG reconstruction pipeline that directly bridges the gap between metagenomics
660 and metabolic modelling, as it proposes add additional pipeline steps for the automatic reconstruction of
661 GENREs using Carveme and metabolic interaction inference using SMETANA. While all-in-one pipelines
662 are valuable, the difficulty of metagenome assembly combined with the complexity of microbial ecosystems
663 maintain the need for curation and additional analyses, especially to capture certain taxonomic clades such
664 as eukaryotes (see subsection 7.2.3).

665 7.2.3 Extending metagenomic assembly to eukaryotes

666 Along with prokaryotes, eukaryotes are significant drivers of microbiomes. For instance, autotrophic and
667 mixotrophic protists fix carbon in aquatic environments, and heterotrophic protists catalyse nutrient cycling
668 in aquatic and terrestrial ecosystems [Worden et al., 2015, Massana and López-Escardó, 2022]. Fungi are
669 much more prevalent soil and aquatic ecosystems [Bahram et al., 2020], and play essential roles in global
670 nutrient cycles and the plant rhizosphere [Tedersoo et al., 2014]. Much can be learned about the diversity
671 and role of eukaryotes in our environment from eukaryotic MAG retrieval, but several technical hurdles pre-
672 vent this currently: (1) eukaryotic genomic complexity, due to larger genomes, multiple chromosomes, and
673 the many non-coding and repetitive regions, complicates both metagenome assembly and MAG retrieval;
674 (2) commonly studied environments, such as the human gut, contain only very few microeukaryotes, at
675 typically low abundance [Bahram et al., 2020, Tito et al., 2019]; and (3) there is a bias in the currently
676 available metagenomic computational tools towards the study of bacterial and archaeal members of the
677 community [Massana and López-Escardó, 2022, Pronk and Medema, 2022].

678 The differences in gene structure between prokaryotes and eukaryotes is a major obstacle to the iden-
679 tification and annotation of eukaryotic genes. Prokaryotic tools applied to eukaryotic contigs tend to
680 predict incomplete, erroneous and discontinuous genes and thus protein sequences. Consequently, eu-
681 karyotic proteins are neglected in microbiome data, with a risk of being truncated and assigned an un-
682 reliable annotations [Pronk and Medema, 2022, Karlicki et al., 2022]. Most public metagenomic repos-

683 itories use annotation pipelines tailored for prokaryotes, regardless of the taxonomic origin of contigs
684 [Lind and Pollard, 2021].

685 Recent methodological developments tackle this challenge, through the identification of eukaryotic con-
686 tigs and genes [Pronk and Medema, 2022, Karlicki et al., 2022, Stanke et al., 2004] and their subsequent
687 annotation, for instance with MetaEuk [Karin et al., 2020]. A few examples of eukaryotic MAGs have
688 been reported, but these relied either on metagenomes with a higher fraction of eukaryotes present (such
689 as e.g. [Duncan et al., 2022]), or on human gut samples that were clearly dysbiotic and had an overgrowth
690 of a single fungal species [Olm et al., 2019]. These examples demonstrate some of the difficulties associated
691 to investigating eukaryotic metagenomes and that methodological challenges remain. Nonetheless, the re-
692 covery of eukaryotic genomes in microbiomes has a major ecological interest and it will be facilitated by
693 TGS.

694 **8. Importance of data reuse and repurposing in metagenomics and** 695 **metabolic modelling**

696 The data pipeline from the reads to the model encompasses a long list of steps, tools and parameters that
697 scientists may use and combine in the process. This raises important questions related to the reproducibility
698 of the analysis outcomes, and the traceability of the data and processes involved. We provide below an
699 overview of the importance of these aspects in the context of the topics treated in the review.

700 Characterising microbial communities requires data, data are acquired in context, and context provides
701 key clues for deciphering the function of a gene or an organism. *Metadata* are descriptions of, and facts
702 about, data and other research outputs¹. Metadata for genomic and metagenomic data from environ-
703 mental samples can describe the context of the source material providing the data, including space, time,
704 habitat, characteristics of the environment, biotic relationships; as well as sequencing and assembly meth-
705 ods [Yilmaz et al., 2011]. For metabolic modelling, metadata can include references to genomes, traces of
706 curation and tools used for reconstruction [Aite et al., 2018]. Indeed, the computational processes that
707 produce research outputs are also performed in a context, which provides key information for reproduc-
708 ing, evaluating, and reusing those research outputs. Metadata can therefore also describe computational
709 processes across an analytic stack consisting of input data, tools, reports, pipelines, and publications

¹<https://schema.datacite.org>

710 [Leipzig et al., 2021].

711 Data and reporting standards must be used in order to guarantee interoperability, reuse, and repurpos-
712 ing. Reuse of publicly available raw sequences of metagenomes, or metagenomic assemblies is very common
713 thanks to dedicated resources such as MGnify [Richardson et al., 2022], and MAGs as well are made avail-
714 able in databases. The Genomic Standards Consortium (GSC) has formalised minimum reporting standards
715 for genomes, metagenomes—Minimum Information about Metagenomic Sequence (MIMS)—and marker
716 gene sequences with these purposes in mind [Yilmaz et al., 2011, ten Hoopen et al., 2017]. In particular,
717 the Minimum Information about a Metagenome-Assembled Genome (MIMAG) [Bowers et al., 2017] is a
718 community-developed standard for reporting bacterial and archaeal genome MAGs.

719 Metabolic modelling is not exempt from considerations related to traceability. Given the variety of tools
720 suitable for GENRE generation, it is important to promote their sharing and standardisation. The use
721 of standardised nomenclatures for GENREs reconstructed in different labs, such as in the BiGG database
722 [Norsigian et al., 2019], is crucial for GENRE comparison and construction of multi-species GEMs. In-
723 deed, a recurring problem lies in the reconciliation of identifiers across databases, for which MetaNetX
724 provides identifier mapping resources [Moretti et al., 2016]. A standard format for the representation and
725 exchange of GENREs but also cell signaling pathways and other systems biology models is the Systems Bi-
726 ology Markup Language [Hucka et al., 2018]. This declarative format based on XML (eXtensible Markup
727 Language) not only permits the representation of the metabolic network, but also includes mathematical
728 parameters used for its simulation. However, the use of this format does not guarantee by itself the quality
729 of the model contents, its reusability with other tools, or the compatibility of the identifiers with different
730 databases. MEMOTE is a community-based test suite for the assessment of GENRE quality, representing a
731 valuable effort in the direction of standardisation, reproducibility and interoperability [Lieven et al., 2020].

732 Widespread adoption of *FAIR* guidelines [Wilkinson et al., 2016] is a further driver of the formal-
733 isation of data and metadata, since the latter are essential for finding and reusing the former. Going
734 beyond this basic requirement, a key opportunity is to promote *reuse* and *repurposing* of research objects
735 [Bechhofer et al., 2013]. Capturing the computational processes that produce research objects, is critical for
736 reproducibility; [Cohen-Boulakia et al., 2017, Leipzig et al., 2021] identify the needs—and challenges—of
737 improved reproducibility and define criteria for “reproducibility-friendly” scientific workflow systems.

738 9. Conclusion and future prospects

739 Improvements in genome reconstruction from metagenomic data have made it possible to build data-
740 informed metabolic models for microbial communities. These mechanistic models describe the phenom-
741 logical behaviour of the community, but furthermore capture the metabolic functions of under-represented
742 species and strain-specific metabolic variations that play an important role in the community dispropor-
743 tionate to their abundance. Third-generation sequencing has proved to be a key technology for answering
744 challenges in metagenomics, providing better recovery of eukaryotic genomes, allowing strain-level analy-
745 sis and pangenomics, and meeting algorithmic challenges in genome delineation. Improved metagenomics
746 leads to improved metabolic models that more fully describe communal functions as well as the added
747 value of individual species and strains. The translation of the latest technologies into routine experimental
748 design will likely take years, however, so algorithmic developments addressing these same questions for
749 NGS short read metagenomics are still needed.

750 Metagenomics is the -omics that contributes the most to community-wide metabolic models, in the
751 same way that genomics contributes to models of individual species. Integrating transcriptomics and
752 metabolomics to build *integrative models*, following a systems biology and systems ecology approach, is
753 necessary to fully comprehend the (eco)system. Even if these data are not obtained systematically, they
754 can be valuable for calibrating metabolic models.

755 Depending on the complexity of the ecosystem and despite the use of TGS, the number of MAGs
756 retrieved from environmental samples is today subpar, and the proportion of raw reads they make use of
757 can remain limited [Pessi et al., 2022]. Those limitations may prevent MAG-derived metabolic models from
758 accurately reflecting the complete functional diversity of the microbiome. Continuing efforts in developing
759 efficient algorithms for MAG reconstruction must be synergetic with sequencing technique improvements.

760 Throughout this review we contrasted phenomenological and mechanistic models as complementary
761 ways of describing the complex behaviour of microbial communities. As alluded to in section 5, hybrid
762 models are a promising avenue for reconciling both approaches. They could combine coarse-grained but
763 highly-scalable phenomenological modelling with limited fine-grained modelling of selected systems, such as
764 the often-overlooked secondary metabolism. Developing a hybrid paradigm would combine the advantages
765 of the two approaches and greatly contribute to building complete understanding of natural microbial
766 communities.

10. Acknowledgements

KC was supported by the MRC Doctoral Antimicrobial Research Training (DART) Industrial CASE Programme. CF, DJS and SL by the French National Research Agency (ANR) France 2030 PEPR Agroécologie et Numérique ANR-22-PEAE-0011 (MISTIC). CF, SL, CM and PUS were supported by the Inria Exploratory Action SLIMMEST. FH was supported by the ERC StG European Research Council H2020 StG (erc-stg-948219, EPYC). FH and KC received support from Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic Program Food Microbiome and Health BB/X011054/1, its constituent project BBS/E/F/000PR13631 as well as Decoding Biodiversity BB/X011089/1, its constituent project(s) BBS/E/T/000PR13638.

11. Conflicts of interest

The authors declare no conflict of interest.

12. Copyright

This research was funded, in whole or in part, by the European Research Council. A CC-BY public copyright license has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission, in accordance with the grant's open access conditions.

References

- [Aite et al., 2018] Aite, M., Chevallier, M., Frioux, C., Trottier, C., Got, J., Cortés, M. P., Mendoza, S. N., Carrier, G., Dameron, O., Guillaudeux, N., Latorre, M., Loira, N., Markov, G. V., Maass, A., and Siegel, A. (2018). Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models. *PLoS Computational Biology*, 14(5):e1006146.
- [Albanese and Donati, 2017] Albanese, D. and Donati, C. (2017). Strain profiling and epidemiology of bacterial species from metagenomic sequencing. *Nature Communications*, 8(1):2260.
- [Alcock et al., 2020] Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A. N., Bordeleau, E., Pawlowski, A. C., Zubyk, H. L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G. L., Beiko, R. G., Brinkman, F. S. L., Hsiao, W. W. L., Domselaar, G. V., and McArthur,

- 793 A. G. (2020). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids*
794 *Research*, 48(D1):D517–D525.
- 795 [Almeida et al., 2019] Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., Lawley, T. D., and Finn,
796 R. D. (2019). A new genomic blueprint of the human gut microbiota. *Nature*, 568(7753):499–504.
- 797 [Almeida et al., 2021] Almeida, A., Nayfach, S., Boland, M., Strozzii, F., Beracochea, M., Shi, Z. J., Pollard, K. S., Sakharova, E., Parks,
798 D. H., Hugenholtz, P., Segata, N., Kyrpides, N. C., and Finn, R. D. (2021). A unified catalog of 204,938 reference genomes from the
799 human gut microbiome. *Nature Biotechnology*, 39(1):105–114.
- 800 [Alneberg et al., 2014] Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J.,
801 Andersson, A. F., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, 11(11):1144–
802 1146.
- 803 [Amarasinghe et al., 2020] Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., and Gouil, Q. (2020). Opportunities and
804 challenges in long-read sequencing data analysis. *Genome Biology*, 21(1):30.
- 805 [Andrews et al., 2010] Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., and Wingett, S. (2010). FastQC.
- 806 [Angeles-Martinez and Hatzimanikatis, 2021] Angeles-Martinez, L. and Hatzimanikatis, V. (2021). Spatio-temporal modeling of the
807 crowding conditions and metabolic variability in microbial communities. *PLOS Computational Biology*, 17(7):e1009140.
- 808 [Arikawa et al., 2021] Arikawa, K., Ide, K., Kogawa, M., Saeki, T., Yoda, T., Endoh, T., Matsushashi, A., Takeyama, H., and Hosokawa,
809 M. (2021). Recovery of strain-resolved genomes from human microbiome through an integration framework of single-cell genomics and
810 metagenomics. *Microbiome*, 9(1):202.
- 811 [Arkin et al., 2018] Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez,
812 F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S.,
813 Drake, M. M., Brettin, T. S., Glass, E. M., Chivian, D., Gunter, D., Weston, D. J., Allen, B. H., Baumohl, J., Best, A. A., Bowen,
814 B., Brenner, S. E., Bun, C. C., Chandonia, J.-M., Chia, J.-M., Colasanti, R., Conrad, N., Davis, J. J., Davison, B. H., DeJongh, M.,
815 Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J. N., Fang, G., Faria, J. P., Frybarger, P. M., Gerlach, W., Gerstein, M., Greiner,
816 A., Gurtowski, J., Haun, H. L., He, F., Jain, R., Joachimiak, M. P., Keegan, K. P., Kondo, S., Kumar, V., Land, M. L., Meyer,
817 F., Mills, M., Novichkov, P. S., Oh, T., Olsen, G. J., Olson, R., Parrello, B., Pasternak, S., Pearson, E., Poon, S. S., Price, G. A.,
818 Ramakrishnan, S., Ranjan, P., Ronald, P. C., Schatz, M. C., Seaver, S. M. D., Shukla, M., Sutormin, R. A., Syed, M. H., Thomason,
819 J., Tintle, N. L., Wang, D., Xia, F., Yoo, H., Yoo, S., and Yu, D. (2018). KBase: The united states department of energy systems
820 biology knowledgebase. *Nature Biotechnology*, 36(7):566–569.
- 821 [Assar et al., 2014] Assar, R., Montecino, M. A., Maass, A., and Sherman, D. J. (2014). Modeling acclimatization by hybrid systems:
822 Condition changes alter biological system behavior models. *Bio Systems*, 121:43–53.
- 823 [Ay et al., 2012] Ay, F., Dang, M., and Kahveci, T. (2012). Metabolic network alignment in large scale by network compression. *BMC*
824 *Bioinformatics*, 13(SUPPL.3):1–19.
- 825 [Ayling et al., 2020] Ayling, M., Clark, M. D., and Leggett, R. M. (2020). New approaches for metagenome assembly with short reads.
826 *Briefings in Bioinformatics*, 21(2):584–594.
- 827 [Babaei et al., 2018] Babaei, P., Shoaie, S., Ji, B., and Nielsen, J. (2018). Challenges in modeling the human gut microbiome. *Nature*
828 *Biotechnology*, 36(8):682–686.

- 829 [Bahram et al., 2018] Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., Bengtsson-
830 Palme, J., Anslan, S., Coelho, L. P., Harend, H., Huerta-Cepas, J., Medema, M. H., Maltz, M. R., Mundra, S., Olsson, P. A., Pent,
831 M., Pöhlme, S., Sunagawa, S., Ryberg, M., Tedersoo, L., and Bork, P. (2018). Structure and function of the global topsoil microbiome.
832 *Nature*, 560(7717):233–237.
- 833 [Bahram et al., 2020] Bahram, M., Netherway, T., Frioux, C., Ferretti, P., Coelho, L. P., Geisen, S., Bork, P., and Hildebrand, F. (2020).
834 Metagenomic assessment of the global distribution of bacteria and fungi. *Environmental Microbiology*.
- 835 [Bartell et al., 2014] Bartell, J. A., Yen, P., Varga, J. J., Goldberg, J. B., and Papin, J. A. (2014). Comparative metabolic systems
836 analysis of pathogenic Burkholderia. *Journal of Bacteriology*, 196(2):210–226.
- 837 [Basile et al., 2020] Basile, A., Campanaro, S., Kovalovszki, A., Zampieri, G., Rossi, A., Angelidaki, I., Valle, G., and Treu, L. (2020).
838 Revealing metabolic mechanisms of interaction in the anaerobic digestion microbiome by flux balance analysis. *Metabolic Engineering*,
839 62:138–149.
- 840 [Basile et al., 2023a] Basile, A., Heinken, A., Hertel, J., Smarr, L., Li, W., Treu, L., Valle, G., Campanaro, S., and Thiele, I. (2023a).
841 Longitudinal flux balance analyses of a patient with episodic colonic inflammation reveals microbiome metabolic dynamics. *Gut*
842 *Microbes*, 15(1):2226921.
- 843 [Basile et al., 2023b] Basile, A., Zampieri, G., Kovalovszki, A., Karkaria, B., Treu, L., Patil, K. R., and Campanaro, S. (2023b). Modelling
844 of microbial interactions in anaerobic digestion: From black to glass box. *Current Opinion in Microbiology*, 75:102363.
- 845 [Batstone et al., 2002] Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, SV., Pavlostathis, SG., Rozzi, A., Sanders, WTM., Siegrist,
846 HA., and Vavilin, VA. (2002). The IWA anaerobic digestion model no 1 (ADM1). *Water Science and technology*, 45(10):65–73.
- 847 [Bauer et al., 2017] Bauer, E., Zimmermann, J., Baldini, F., Thiele, I., and Kaleta, C. (2017). BacArena: Individual-based metabolic
848 modeling of heterogeneous microbes in complex communities. *PLOS Computational Biology*, 13(5):e1005544.
- 849 [Bechhofer et al., 2013] Bechhofer, S., Buchan, I., Roure, D. D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D.,
850 Delderfield, M., Dunlop, I., Gamble, M., Michaelides, D., Owen, S., Newman, D., Sufi, S., and Goble, C. (2013). Why linked data is
851 not enough for scientists. *Future Generation Computer Systems*, 29(2):599–611.
- 852 [Beghini et al., 2021] Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P.,
853 Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., and Segata,
854 N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife*,
855 10:e65088.
- 856 [Belcour et al., 2020] Belcour, A., Frioux, C., Aite, M., Bretaudeau, A., Hildebrand, F., and Siegel, A. (2020). Metage2Metabo,
857 microbiota-scale metabolic complementarity for the identification of key species. *eLife*, 9:e61968.
- 858 [Belcour et al., 2023] Belcour, A., Got, J., Aite, M., Delage, L., Collén, J., Frioux, C., Leblanc, C., Dittami, S. M., Blanquart, S., Markov,
859 G. V., and Siegel, A. (2023). Inferring and comparing metabolism across heterogeneous sets of annotated genomes using AuCoMe.
860 *Genome Research*.
- 861 [Benoit et al., 2023] Benoit, G., Raguideau, S., James, R., Phillippy, A. M., Chikhi, R., and Quince, C. (2023). Efficient high-quality
862 metagenome assembly from long accurate reads using minimizer-space de bruijn graphs. *bioRxiv : the preprint server for biology*.
- 863 [Bernardini et al., 2022] Bernardini, N. D., Basile, A., Zampieri, G., Kovalovszki, A., Diaz, B. D. D., Offer, E., Wongfaed, N., Angelidaki,
864 I., Kougias, P. G., Campanaro, S., and Treu, L. (2022). Integrating metagenomic binning with flux balance analysis to unravel
865 syntrophies in anaerobic CO₂ methanation. *Microbiome*, 10(1):117.

- 866 [Bernstein et al., 2019] Bernstein, D. B., Dewhirst, F. E., and Segrè, D. (2019). Metabolic network percolation quantifies biosynthetic
867 capabilities across the human oral microbiome. *eLife*, 8:e39733.
- 868 [Bernstein et al., 2021] Bernstein, D. B., Sulheim, S., Almaas, E., and Segrè, D. (2021). Addressing uncertainty in genome-scale metabolic
869 model reconstruction and analysis. *Genome Biology*, 22(1):64.
- 870 [Bertrand et al., 2019] Bertrand, D., Shaw, J., Kalathiyappan, M., Ng, A. H. Q., Kumar, M. S., Li, C., Dvornicic, M., Soldo, J. P., Koh,
871 J. Y., Tong, C., Ng, O. T., Barkham, T., Young, B., Marimuthu, K., Chng, K. R., Sikic, M., and Nagarajan, N. (2019). Hybrid
872 metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nature*
873 *Biotechnology*, 37(8):937–944.
- 874 [Bickhart et al., 2022] Bickhart, D. M., Kolmogorov, M., Tseng, E., Portik, D. M., Korobeynikov, A., Tolstoganov, I., Uritskiy, G.,
875 Liachko, I., Sullivan, S. T., Shin, S. B., Zorea, A., Andreu, V. P., Panke-Buisse, K., Medema, M. H., Mizrahi, I., Pevzner, P. A., and
876 Smith, T. P. L. (2022). Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities.
877 *Nature Biotechnology*, 40(5):711–719.
- 878 [Biggs et al., 2015] Biggs, M. B., Medlock, G. L., Kolling, G. L., and Papin, J. A. (2015). Metabolic network modeling of microbial
879 communities. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 7(5):317–334.
- 880 [Blanco-Miguez et al., 2022] Blanco-Miguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois,
881 L., Huang, K. D., Thomas, A. M., Piccinno, G., Piperni, E., Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R.,
882 Wolf, J., Berry, S. E., Spector, T. D., Franzosa, E. A., Pasolli, E., Asnicar, F., Huttenhower, C., and Segata, N. (2022). Extending
883 and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4. *bioRxiv : the preprint server for*
884 *biology*, page 2022.08.22.504593.
- 885 [Blin et al., 2020] Blin, K., Shaw, S., Kautsar, S. A., Medema, M. H., and Weber, T. (2020). The antiSMASH database version 3:
886 Increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Research*, 49(D1):gkaa978–.
- 887 [Bolger et al., 2014] Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data.
888 *Bioinformatics (Oxford, England)*, 30(15):2114–2120.
- 889 [Borenstein and Feldman, 2009] Borenstein, E. and Feldman, M. W. (2009). Topological signatures of species interactions in metabolic
890 networks. <https://home.liebertpub.com/cmb>, 16(2):191–200.
- 891 [Borenstein et al., 2008] Borenstein, E., Kupiec, M., Feldman, M. W., and Ruppin, E. (2008). Large-scale reconstruction and phylogenetic
892 analysis of metabolic environments. *Proceedings of the National Academy of Sciences*, 105(38):14482–14487.
- 893 [Borer et al., 2019] Borer, B., Ataman, M., Hatzimanikatis, V., and Or, D. (2019). Modeling metabolic networks of individual bacterial
894 agents in heterogeneous and dynamic soil habitats (IndiMeSH). *PLOS Computational Biology*, 15(6):e1007127.
- 895 [Bosi et al., 2016] Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., and Palsson, B. (2016). Comparative genome-scale modelling
896 of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National*
897 *Academy of Sciences of the United States of America*, 113(26):E3801–E3809.
- 898 [Bowers et al., 2017] Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F.,
899 Jarett, J., Rivers, A. R., Eloie-Fadrosch, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom,
900 R. R., Birren, B., Podar, M., Bork, P., Weinstock, G. M., Garrity, G. M., Dodsworth, J. A., Yooseph, S., Sutton, G., Glöckner, F. O.,
901 Gilbert, J. A., Nelson, W. C., Hallam, S. J., Jungbluth, S. P., Ettema, T. J. G., Tighe, S., Konstantinidis, K. T., Liu, W.-T., Baker,

- 902 B. J., Rattei, T., Eisen, J. A., Hedlund, B., McMahon, K. D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I.,
903 Tyson, G. W., Rinke, C., Kyrpides, N. C., Schriml, L., Garrity, G. M., Hugenholtz, P., Sutton, G., Yilmaz, P., Meyer, F., Glöckner,
904 F. O., Gilbert, J. A., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D. H., Eren,
905 A. M., Schriml, L., Banfield, J. F., Hugenholtz, P., and Woyke, T. (2017). Minimum information about a single amplified genome
906 (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, 35(8):725–731.
- 907 [Bowers et al., 2022] Bowers, R. M., Nayfach, S., Schulz, F., Jungbluth, S. P., Ruhl, I. A., Sheremet, A., Lee, J., Goudeau, D., Eloe-
908 Fadrosch, E. A., Stepanauskas, R., Malmstrom, R. R., Kyrpides, N. C., Dunfield, P. F., and Woyke, T. (2022). Dissecting the
909 dominant hot spring microbial populations based on community-wide sampling at single-cell genomic resolution. *The ISME Journal*,
910 16(5):1337–1347.
- 911 [Bowman and Ducklow, 2015] Bowman, J. S. and Ducklow, H. W. (2015). Microbial communities can be described by metabolic structure:
912 A general framework and application to a seasonally variable, depth-stratified microbial community from the coastal west antarctic
913 peninsula. *PLoS ONE*, 10(8):e0135868.
- 914 [Bradley et al., 2016] Bradley, I. M., Pinto, A. J., and Guest, J. S. (2016). Design and evaluation of illumina MiSeq-Compatible, 18S rRNA
915 gene-specific primers for improved characterization of mixed phototrophic communities. *Applied and Environmental Microbiology*,
916 82(19):5878–5891.
- 917 [Branicky, 1994] Branicky, M. (1994). Stability of switched and hybrid systems. *Proceedings of 1994 33rd IEEE Conference on Decision
918 and Control*, 4:3498–3503 vol.4.
- 919 [Brown et al., 2021] Brown, C. L., Keenum, I. M., Dai, D., Zhang, L., Vikesland, P. J., and Pruden, A. (2021). Critical evaluation
920 of short, long, and hybrid assembly for contextual analysis of antibiotic resistance genes in complex environmental metagenomes.
921 *Scientific Reports*, 11(1):3753.
- 922 [Buchfink et al., 2014] Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature
923 Methods*, 12(1):59–60.
- 924 [Budnich et al., 2017] Budnich, M., Bourdon, J., Larhlimi, A., and Eveillard, D. (2017). A multi-objective constraint-based approach
925 for modeling genome-scale microbial ecosystems. *PLOS ONE*, 12(2):e0171744.
- 926 [Bukin et al., 2019] Bukin, Y. S., Galachyants, Y. P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., and Zemskaya, T. I. (2019). The
927 effect of 16S rRNA region choice on bacterial community metabarcoding results. *Scientific Data*, 6(1):190007.
- 928 [Burki et al., 2020] Burki, F., Roger, A. J., Brown, M. W., and Simpson, A. G. (2020). The new tree of eukaryotes. *Trends in Ecology
929 & Evolution*, 35(1):43–55.
- 930 [Callahan et al., 2016] Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2:
931 High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7):581–583.
- 932 [Cantalapiedra et al., 2021] Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-
933 mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and
934 Evolution*, 38(12):msab293–.
- 935 [Cao et al., 2016] Cao, Y., Wang, Y., Zheng, X., Li, F., and Bo, X. (2016). RevEcoR: An R package for the reverse ecology analysis of
936 microbiomes. *BMC Bioinformatics*, 17(1):1–6.

- 937 [Caspi et al., 2019] Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., Ong, W. K., Paley, S.,
938 Subhraveti, P., and Karp, P. D. (2019). The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids*
939 *Research*, 48(D1):D445–D453.
- 940 [Chang et al., 2020] Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., and
941 Schomburg, D. (2020). BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Research*,
942 49(D1):D498–D508.
- 943 [Chen et al., 2020] Chen, L.-X., Anantharaman, K., Shaiber, A., Eren, A. M., and Banfield, J. F. (2020). Accurate and complete genomes
944 from metagenomes. *Genome Research*, 30(3):315–333.
- 945 [Christian et al., 2007] Christian, N., Handorf, T., and Ebenhöf, O. (2007). Metabolic synergy: Increasing biosynthetic capabilities by
946 network cooperation. *Genome informatics International Conference on Genome Informatics*, 18:320–329.
- 947 [Churchward et al., 2022] Churchward, B., Millet, M., Bihouée, A., Fertin, G., and Chaffron, S. (2022). MAGNETO: An automated
948 workflow for genome-resolved metagenomics. *mSystems*, 7(4):e00432–22.
- 949 [Cohen-Boulakia et al., 2017] Cohen-Boulakia, S., Belhajjame, K., Collin, O., Chopard, J., Froidevaux, C., Gaignard, A., Hinsén, K.,
950 Larmande, P., Bras, Y. L., Lemoine, F., Mareuil, F., Ménager, H., Pradal, C., and Blanchet, C. (2017). Scientific workflows for
951 computational reproducibility in the life sciences: Status, challenges and opportunities. *Future Generation Computer Systems*, 75:284–
952 298.
- 953 [Colarusso et al., 2021] Colarusso, A. V., Goodchild-Michelman, I., Rayle, M., and Zomorodi, A. R. (2021). Computational modeling
954 of metabolism in microbial communities on a genome-scale. *Current Opinion in Systems Biology*.
- 955 [Coleman et al., 2007] Coleman, M. C., Fish, R., and Block, D. E. (2007). Temperature-dependent kinetic model for nitrogen-limited
956 wine fermentations. *Applied and Environmental Microbiology*, 73(18):5875–5884.
- 957 [Commichaux et al., 2021] Commichaux, S., Shah, N., Ghurye, J., Stoppel, A., Goodheart, J. A., Luque, G. G., Cummings, M. P., and
958 Pop, M. (2021). A critical assessment of gene catalogs for metagenomic analysis. *Bioinformatics (Oxford, England)*, 37(18):2848–2857.
- 959 [Consortium et al., 2021] Consortium, T. U., Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi,
960 E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. D., Denny, P., Dogan, T.,
961 Ebenezer, T., Fan, J., Castro, L. G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana,
962 G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F.,
963 Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C. S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez,
964 M. R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins,
965 X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimò, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P.,
966 Baratin, D., Blatter, M.-C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K. C., Coudert, E.,
967 Cuhe, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M. L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos,
968 A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Mercier, P. L.,
969 Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T. B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L.,
970 Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C. H., Arighi,
971 C. N., Arminski, L., Chen, C., Chen, Y., Garavelli, J. S., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka,
972 C. R., Wang, Q., Wang, Y., Yeh, L.-S., Zhang, J., Ruch, P., and Teodoro, D. (2021). UniProt: The universal protein knowledgebase
973 in 2021. *Nucleic Acids Research*, 49(D1):D480–D489.

- 974 [Costa et al., 2016] Costa, R. S., Hartmann, A., and Vinga, S. (2016). Kinetic modeling of cell metabolism for microbial production.
975 *Journal of Biotechnology*, 219:126–141.
- 976 [Coster et al., 2018] Coster, W. D., D’Hert, S., Schultz, D. T., Cruts, M., and Broeckhoven, C. V. (2018). NanoPack: Visualizing and
977 processing long-read sequencing data. *Bioinformatics (Oxford, England)*, 34(15):2666–2669.
- 978 [Coster and Rademakers, 2023] Coster, W. D. and Rademakers, R. (2023). NanoPack2: Population-scale evaluation of long-read se-
979 quencing data. *Bioinformatics (Oxford, England)*, 39(5):btad311.
- 980 [Coyte and Rakoff-Nahoum, 2019] Coyte, K. Z. and Rakoff-Nahoum, S. (2019). Understanding competition and cooperation within the
981 mammalian gut microbiome. *Current Biology*, 29(11):R538–R544.
- 982 [Darrigade et al., 2022] Darrigade, L., Haghebaert, M., Cherbuy, C., Labarthe, S., and Laroche, B. (2022). A PDMP model of the
983 epithelial cell turn-over in the intestinal crypt including microbiota-derived regulations. *Journal of Mathematical Biology*, 84(7):60.
- 984 [de Muinck et al., 2017] de Muinck, E. J., Trosvik, P., Gilfillan, G. D., Hov, J. R., and Sundaram, A. Y. M. (2017). A novel ultra
985 high-throughput 16S rRNA gene amplicon sequencing library preparation method for the Illumina HiSeq platform. *Microbiome*,
986 5(1):68.
- 987 [de Oliveira Martins et al., 2019] de Oliveira Martins, L., Page, A. J., Mather, A. E., and Charles, I. G. (2019). Taxonomic resolution
988 of the ribosomal RNA operon in bacteria: Implications for its use with long-read sequencing. *NAR Genomics and Bioinformatics*,
989 2(1):lqz016.
- 990 [DeAngelis and Grimm, 2014] DeAngelis, D. L. and Grimm, V. (2014). Individual-based models in ecology after four decades. *F1000prime*
991 *reports*, 6.
- 992 [Delahaye and Nicolas, 2021] Delahaye, C. and Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PLOS ONE*,
993 16(10):e0257521.
- 994 [Desmond-Le Quémener et al., 2021] Desmond-Le Quémener, E., Moscoviz, R., Bernet, N., and Marcus, A. (2021). Modeling of inter-
995 species electron transfer in anaerobic microbial communities. *Current Opinion in Biotechnology*, 67:49–57.
- 996 [Diener et al., 2020] Diener, C., Gibbons, S. M., and Resendis-Antonio, O. (2020). MICOM: Metagenome-Scale Modeling To Infer
997 Metabolic Interactions in the Gut Microbiota. *mSystems*, 5(1):e00606–19.
- 998 [Djemiel et al., 2022] Djemiel, C., Maron, P.-A., Terrat, S., Dequiedt, S., Cottin, A., and Ranjard, L. (2022). Inferring microbiota
999 functions from taxonomic genes: A review. *GigaScience*, 11:giab090.
- 1000 [Douglas et al., 2020] Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., and
1001 Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, 38(6):685–688.
- 1002 [Drula et al., 2022] Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The carbohydrate-active
1003 enzyme database: Functions and literature. *Nucleic Acids Research*, 50(D1):D571–D577.
- 1004 [Dueholm et al., 2022] Dueholm, M. K. D., Nierychlo, M., Andersen, K. S., Rudkjøbing, V., Knutsson, S., Albertsen, M., and Nielsen,
1005 P. H. (2022). MiDAS 4: A global catalogue of full-length 16S rRNA gene sequences and taxonomy for studies of bacterial communities
1006 in wastewater treatment plants. *Nature communications*, 13(1):1908.

- 1007 [Dukovski et al., 2021] Dukovski, I., Bajić, D., Chacón, J. M., Quintin, M., Vila, J. C. C., Sulheim, S., Pacheco, A. R., Bernstein, D. B.,
1008 Riehl, W. J., Korolev, K. S., Sanchez, A., Harcombe, W. R., and Segrè, D. (2021). A metabolic modeling platform for the computation
1009 of microbial ecosystems in time and space (COMETS). *Nature Protocols*, 16(11):5030–5082.
- 1010 [Duncan et al., 2022] Duncan, A., Barry, K., Daum, C., Eloë-Fadrosh, E., Roux, S., Schmidt, K., Tringe, S. G., Valentin, K. U., Varghese,
1011 N., Salamov, A., Grigoriev, I. V., Leggett, R. M., Moulton, V., and Mock, T. (2022). Metagenome-assembled genomes of phytoplankton
1012 microbiomes from the arctic and atlantic oceans. *Microbiome*, 10(67).
- 1013 [Earl et al., 2018] Earl, J. P., Adappa, N. D., Krol, J., Bhat, A. S., Balashov, S., Ehrlich, R. L., Palmer, J. N., Workman, A. D., Blasetti,
1014 M., Sen, B., Hammond, J., Cohen, N. A., Ehrlich, G. D., and Mell, J. C. (2018). Species-level bacterial community profiling of the
1015 healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes. *Microbiome*, 6(1):190.
- 1016 [Ebenhöh et al., 2004] Ebenhöh, O., Handorf, T., and Heinrich, R. (2004). Structural analysis of expanding metabolic networks. *Genome
1017 informatics. International Conference on Genome Informatics*, 15(1):35–45.
- 1018 [Ebrahim et al., 2013] Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). COBRApy: COstraints-Based Recon-
1019 struction and Analysis for Python. *BMC Systems Biology*, 7(1):74.
- 1020 [Eisenhofer et al., 2023] Eisenhofer, R., Odriozola, I., and Alberdi, A. (2023). Impact of microbial genome completeness on metagenomic
1021 functional inference. *ISME Communications*, 3(1):12.
- 1022 [Eren et al., 2021] Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., Fink, I., Pan, J. N., Yousef, M., Fogarty,
1023 E. C., Trigodet, F., Watson, A. R., Esen, Ö. C., Moore, R. M., Clayssen, Q., Lee, M. D., Kivenson, V., Graham, E. D., Merrill, B. D.,
1024 Karkman, A., Blankenberg, D., Eppley, J. M., Sjödin, A., Scott, J. J., Vázquez-Campos, X., McKay, L. J., McDaniel, E. A., Stevens,
1025 S. L. R., Anderson, R. E., Fuessel, J., Fernandez-Guerra, A., Maignien, L., Delmont, T. O., and Willis, A. D. (2021). Community-led,
1026 integrated, reproducible multi-omics with anvi'o. *Nature Microbiology*, 6(1):3–6.
- 1027 [Estaki et al., 2020] Estaki, M., Jiang, L., Bokulich, N. A., McDonald, D., González, A., Kosciulek, T., Martino, C., Zhu, Q., Birmingham,
1028 A., Vázquez-Baeza, Y., Dillon, M. R., Bolyen, E., Caporaso, J. G., and Knight, R. (2020). QIIME 2 enables comprehensive End-to-End
1029 analysis of diverse microbiome data and comparative studies with publicly available data. *Current Protocols in Bioinformatics*, 70(1).
- 1030 [Faust and Raes, 2012] Faust, K. and Raes, J. (2012). Microbial interactions: From networks to models. *Nature Reviews Microbiology*,
1031 10(8):538–550.
- 1032 [Feist and Palsson, 2010] Feist, A. M. and Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology*,
1033 13(3):344–349.
- 1034 [Feng et al., 2022] Feng, X., Cheng, H., Portik, D., and Li, H. (2022). Metagenome assembly of high-fidelity long reads with hifiasm-meta.
1035 *Nature Methods*, 19(6):671–674.
- 1036 [Fischer et al., 2017] Fischer, M., Strauch, B., and Renard, B. Y. (2017). Abundance estimation and differential testing on strain level
1037 in metagenomics data. *Bioinformatics (Oxford, England)*, 33(14):i124–i132.
- 1038 [Flissi et al., 2020] Flissi, A., Ricart, E., Campart, C., Chevalier, M., Dufresne, Y., Michalik, J., Jacques, P., Flahaut, C., Lisacek, F.,
1039 Leclère, V., and Pupin, M. (2020). Norine: Update of the nonribosomal peptide resource. *Nucleic Acids Research*, 48(D1):D465–D469.
- 1040 [Foundation et al., 2022] Foundation, T. O., Abreu, A., Bourgois, E., Gristwood, A., Troublé, R., Oceans, T., Acinas, S. G., Bork, P.,
1041 Boss, E., Bowler, C., Budinich, M., Chaffron, S., de Vargas, C., Delmont, T. O., Eveillard, D., Guidi, L., Iudicone, D., Kandels,
1042 S., Morlon, H., Lombard, F., Pepperkok, R., Karlusich, J. J. P., Piganeau, G., Régimbeau, A., Sommeria-Klein, G., Stemann, L.,

- 1043 Sullivan, M. B., Sunagawa, S., Wincker, P., Zablocki, O., (EMBL), E. M. B. L., Arendt, D., Bilic, J., Finn, R., Heard, E., Rouse, B.,
1044 Vamathevan, J., (EMBRIC-ERIC), E. M. B. R. C. E. R. I. C., Casotti, R., Cancio, I., Cunliffe, M., Kervella, A. E., Kooistra, W. H.
1045 C. F., Obst, M., Pade, N., Power, D. M., Santi, I., Tsagaraki, T. M., and Vanaverbeke, J. (2022). Priorities for ocean microbiome
1046 research. *Nature Microbiology*, 7(7):937–947.
- 1047 [Franzosa et al., 2018] Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., Lipson, K. S., Knight,
1048 R., Caporaso, J. G., Segata, N., and Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes.
1049 *Nature Methods*, 15(11):962–968.
- 1050 [Freilich et al., 2011] Freilich, S., Zarecki, R., Eilam, O., Segal, E. S., Henry, C. S., Kupiec, M., Gophna, U., Sharan, R., and Ruppin, E.
1051 (2011). Competitive and cooperative metabolic interactions in bacterial communities. *Nature Communications* 2011 2:1, 2(1):1–7.
- 1052 [Frioux et al., 2018] Frioux, C., Fremy, E., Trottier, C., and Siegel, A. (2018). Scalable and exhaustive screening of metabolic functions
1053 carried out by microbial consortia. *Bioinformatics (Oxford, England)*, 34(17):i934–i943.
- 1054 [Frioux et al., 2020] Frioux, C., Singh, D., Korcsmaros, T., and Hildebrand, F. (2020). From bag-of-genes to bag-of-genomes: Metabolic
1055 modelling of communities in the era of metagenome-assembled genomes. *Computational and Structural Biotechnology Journal*, 18:1722–
1056 1734.
- 1057 [Fu et al., 2012] Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing
1058 data. *Bioinformatics (Oxford, England)*, 28(23):3150–3152.
- 1059 [Fukasawa et al., 2020] Fukasawa, Y., Ermini, L., Wang, H., Carty, K., and Cheung, M.-S. (2020). LongQC: A quality control tool for
1060 third generation sequencing long read data. *G3: Genes/Genomes/Genetics*, 10(4):1193–1196.
- 1061 [Galperin et al., 2020] Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., and Koonin, E. V. (2020). COG
1062 database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research*, 49(D1):gkaa1018–
1063 .
- 1064 [García-Jiménez et al., 2018] García-Jiménez, B., García, J. L., and Nogales, J. (2018). FLYCOP: Metabolic modeling-based analysis
1065 and engineering microbial communities. *Bioinformatics (Oxford, England)*, 34(17):i954–i963.
- 1066 [García-Jiménez et al., 2021] García-Jiménez, B., Torres-Bacete, J., and Nogales, J. (2021). Metabolic modelling approaches for describ-
1067 ing and engineering microbial communities. *Computational and Structural Biotechnology Journal*, 19:226–246.
- 1068 [Geng et al., 2021] Geng, J., Ji, B., Li, G., López-Isunza, F., and Nielsen, J. (2021). CODY enables quantitatively spatiotemporal
1069 predictions on in vivo gut microbial variability induced by diet intervention. *Proceedings of the National Academy of Sciences*,
1070 118(13):e2019336118.
- 1071 [Goelzer et al., 2015] Goelzer, A., Muntel, J., Chubukov, V., Jules, M., Prestel, E., Nölker, R., Mariadassou, M., Aymerich, S., Hecker,
1072 M., Noirot, P., Becher, D., and Fromion, V. (2015). Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic*
1073 *Engineering*, 32:232–243.
- 1074 [Gottstein et al., 2016] Gottstein, W., Olivier, B. G., Bruggeman, F. J., and Teusink, B. (2016). Constraint-based stoichiometric mod-
1075 elling from single organisms to microbial communities. *Journal of The Royal Society Interface*, 13(124):20160627.
- 1076 [Grant et al., 2023] Grant, A., Aleidan, A., Davies, C. S., Udochi, S. C., Fritscher, J., Bahram, M., and Hildebrand, F. (2023). Improved
1077 taxonomic annotation of Archaea communities using LotuS2, the Genome Taxonomy Database and RNAseq data. *bioRxiv*.

- 1078 [Greenblum et al., 2012] Greenblum, S., Turnbaugh, P. J., and Borenstein, E. (2012). Metagenomic systems biology of the human gut
1079 microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proceedings of the National Academy
1080 of Sciences of the United States of America*, 109(2):594–599.
- 1081 [Gu et al., 2019] Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., and Lee, S. Y. (2019). Current status and applications of genome-scale
1082 metabolic models. *Genome Biology*, 20(1):121.
- 1083 [Gurbich et al., 2023] Gurbich, T. A., Almeida, A., Beracochea, M., Burdett, T., Burgin, J., Cochrane, G., Raj, S., Richardson, L.,
1084 Rogers, A. B., Sakharova, E., Salazar, G. A., and Finn, R. D. (2023). MGnify Genomes: A resource for biome-specific microbial
1085 genome catalogues. *Journal of Molecular Biology*, page 168016.
- 1086 [Hakimzadeh et al., 2023] Hakimzadeh, A., Asbun, A. A., Albanese, D., Bernard, M., Buchner, D., Callahan, B., Caporaso, J. G., Curd,
1087 E., Djemiel, C., Durling, M. B., Elbrecht, V., Gold, Z., Gweon, H. S., Hajibabaei, M., Hildebrand, F., Mikryukov, V., Normandeau, E.,
1088 Özkurt, E., Palmer, J. M., Pascal, G., Porter, T. M., Straub, D., Vasar, M., Větrovský, T., Zafeiropoulos, H., and Anslan, S. (2023).
1089 A pile of pipelines: An overview of the bioinformatics software for metabarcoding data analyses. *Molecular Ecology Resources*.
- 1090 [Hamilton and Reed, 2012] Hamilton, J. J. and Reed, J. L. (2012). Identification of functional differences in metabolic networks using
1091 comparative genomics and constraint-based models. *PLoS ONE*, 7(4).
- 1092 [Heinken et al., 2023] Heinken, A., Hertel, J., Acharya, G., Ravcheev, D. A., Nyga, M., Okpala, O. E., Hogan, M., Magnúsdóttir, S.,
1093 Martinelli, F., Nap, B., Preciat, G., Edirisinghe, J. N., Henry, C. S., Fleming, R. M. T., and Thiele, I. (2023). Genome-scale metabolic
1094 reconstruction of 7,302 human microorganisms for personalized medicine. *Nature Biotechnology*, pages 1–12.
- 1095 [Heinken et al., 2013] Heinken, A., Sahoo, S., Fleming, R. M. T., and Thiele, I. (2013). Systems-level characterization of a host-microbe
1096 metabolic symbiosis in the mammalian gut. *Gut Microbes*, 4(1):28–40.
- 1097 [Heinken and Thiele, 2015] Heinken, A. and Thiele, I. (2015). Anoxic Conditions Promote Species-Specific Mutualism between Gut
1098 Microbes *In Silico*. *Applied and Environmental Microbiology*, 81(12):4049–4061.
- 1099 [Heinken and Thiele, 2022a] Heinken, A. and Thiele, I. (2022a). Microbiome Modelling Toolbox 2.0: Efficient, tractable modelling of
1100 microbiome communities. *Bioinformatics (Oxford, England)*, 38(8):2367–2368.
- 1101 [Heinken and Thiele, 2022b] Heinken, A. and Thiele, I. (2022b). Microbiome Modelling Toolbox 2.0: Efficient, tractable modelling of
1102 microbiome communities. *Bioinformatics (Oxford, England)*.
- 1103 [Heintz-Buschart and Wilmes, 2018] Heintz-Buschart, A. and Wilmes, P. (2018). Human gut microbiome: Function matters. *Trends in
1104 Microbiology*, 26(7):563–574.
- 1105 [Heirendt et al., 2019] Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak,
1106 J., Keating, S. M., Vlasov, V., Magnúsdóttir, S., Ng, C. Y., Preciat, G., Žagare, A., Chan, S. H. J., Aurich, M. K., Clancy, C. M.,
1107 Modamio, J., Sauls, J. T., Noronha, A., Bordbar, A., Cousins, B., Assal, D. C. E., Valcarcel, L. V., Apaolaza, I., Ghaderi, S.,
1108 Ahookhosh, M., Guebila, M. B., Kostromins, A., Sompairac, N., Le, H. M., Ma, D., Sun, Y., Wang, L., Yurkovich, J. T., Oliveira,
1109 M. A. P., Vuong, P. T., Assal, L. P. E., Kuperstein, I., Zinovyev, A., Hinton, H. S., Bryant, W. A., Artacho, F. J. A., Planes, F. J.,
1110 Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M. A., Maranas, C. D., Lewis, N. E., Sauter, T., Palsson, B. O., Thiele,
1111 I., and Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0.
1112 *Nature Protocols*, 14(3):639–702.
- 1113 [Hellweger et al., 2016] Hellweger, F. L., Clegg, R. J., Clark, J. R., Plugge, C. M., and Kreft, J.-U. (2016). Advancing microbial sciences
1114 by individual-based modelling. *Nature Reviews Microbiology*, 14(7):461–471.

- 1115 [Henry et al., 2016] Henry, C. S., Bernstein, H. C., Weisenhorn, P., Taylor, R. C., Lee, J.-Y., Zucker, J., and Song, H.-S. (2016).
1116 Microbial Community Metabolic Modeling: A Community Data-Driven Network Reconstruction. *Journal of Cellular Physiology*,
1117 231(11):2339–2345.
- 1118 [Henry et al., 2007] Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2007). Thermodynamics-Based Metabolic Flux Analysis.
1119 *Biophysical Journal*, 92(5):1792–1805.
- 1120 [Henzinger, 1996] Henzinger, T. (1996). The theory of hybrid automata. *Proceedings 11th Annual IEEE Symposium on Logic in Computer
1121 Science*, pages 278–292.
- 1122 [Hildebrand, 2021] Hildebrand, F. (2021). Ultra-resolution metagenomics: When enough is not enough. *mSystems*, 6(4):e00881–21.
- 1123 [Hildebrand et al., 2021] Hildebrand, F., Gossmann, T. I., Frioux, C., Özkurt, E., Myers, P. N., Ferretti, P., Kuhn, M., Bahram, M.,
1124 Nielsen, H. B., and Bork, P. (2021). Dispersal strategies shape persistence and evolution of human gut bacteria. *Cell Host & Microbe*.
- 1125 [Hildebrand et al., 2019] Hildebrand, F., Moitinho-Silva, L., Blasche, S., Jahn, M. T., Gossmann, T. I., Huerta-Cepas, J., Hercog, R.,
1126 Luetge, M., Bahram, M., Pryszyk, A., Alves, R. J., Waszak, S. M., Zhu, A., Ye, L., Costea, P. I., Aalvink, S., Belzer, C., Forslund,
1127 S. K., Sunagawa, S., Hentschel, U., Merten, C., Patil, K. R., Benes, V., and Bork, P. (2019). Antibiotics-induced monodominance of
1128 a novel gut bacterial order. *Gut*, 68(10):1781.
- 1129 [Hildebrand et al., 2014] Hildebrand, F., Tadeo, R., Voigt, A. Y., Bork, P., and Raes, J. (2014). LotuS: An efficient and user-friendly
1130 OTU processing pipeline. *Microbiome*, 2(1):30.
- 1131 [Hrovatin et al., 2021] Hrovatin, K., Fischer, D. S., and Theis, F. J. (2021). Towards modelling metabolic state from single-cell tran-
1132 scriptomics. *Molecular Metabolism*, page 101396.
- 1133 [Hucka et al., 2018] Hucka, M., Bergmann, F. T., Dräger, A., Hoops, S., Keating, S. M., Novère, N. L., Myers, C. J., Olivier, B. G.,
1134 Sahle, S., Schaff, J. C., Smith, L. P., Waltemath, D., and Wilkinson, D. J. (2018). The systems biology markup language (SBML):
1135 Language specification for level 3 version 2 core. *Journal of Integrative Bioinformatics*, 15(1):20170081.
- 1136 [Huerta-Cepas et al., 2018] Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R.,
1137 Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., and Bork, P. (2018). eggNOG 5.0: A hierarchical, functionally and phylogenetically
1138 annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1):gky1085.
- 1139 [Hyatt et al., 2010] Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: Prokaryotic
1140 gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1):119.
- 1141 [Jamshidi and Palsson, 2010] Jamshidi, N. and Palsson, B. Ø. (2010). Mass Action Stoichiometric Simulation Models: Incorporating
1142 Kinetics and Regulation into Stoichiometric Models. *Biophysical Journal*, 98(2):175–185.
- 1143 [Jenior et al., 2020] Jenior, M. L., Moutinho, T. J., Dougherty, B. V., and Papin, J. A. (2020). Transcriptome-guided parsimonious flux
1144 analysis improves predictions with metabolic networks in complex environments. *PLOS Computational Biology*, 16(4):e1007099.
- 1145 [Jing et al., 2021] Jing, G., Zhang, Y., Cui, W., Liu, L., Xu, J., and Su, X. (2021). Meta-Apo improves accuracy of 16S-amplicon-based
1146 prediction of microbiome function. *BMC Genomics*, 22(1):9.
- 1147 [Johnson et al., 2019] Johnson, J. S., Spakowicz, D. J., Hong, B.-Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson,
1148 B. M., Agresta, H. O., Gerstein, M., Sodergren, E., and Weinstock, G. M. (2019). Evaluation of 16S rRNA gene sequencing for species
1149 and strain-level microbiome analysis. *Nature Communications*, 10:5029.

- 1150 [Kanehisa et al., 2021] Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: Integrating
1151 viruses and cellular organisms. *Nucleic Acids Research*, 49(D1):gkaa970–.
- 1152 [Kang et al., 2019] Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: An adaptive
1153 binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7:e7359.
- 1154 [Karimi et al., 2021] Karimi, E., Geslain, E., Belcour, A., Frioux, C., Aïte, M., Siegel, A., Corre, E., and Dittami, S. M. (2021).
1155 Robustness analysis of metabolic predictions in algal microbial communities based on different annotation pipelines. *PeerJ*, 9:e11344.
- 1156 [Karimian and Motamedian, 2020] Karimian, E. and Motamedian, E. (2020). ACBM: An Integrated Agent and Constraint Based Mod-
1157 eling Framework for Simulation of Microbial Communities. *Scientific Reports*, 10(1):8695.
- 1158 [Karin et al., 2020] Karin, E. L., Mirdita, M., and Söding, J. (2020). MetaEuk—sensitive, high-throughput gene discovery, and annotation
1159 for large-scale eukaryotic metagenomics. *Microbiome*, 8(1):48.
- 1160 [Karlicki et al., 2022] Karlicki, M., Antonowicz, S., and Karnkowska, A. (2022). Tiara: Deep learning-based classification system for
1161 eukaryotic sequences. *Bioinformatics (Oxford, England)*, 38(2):344–350.
- 1162 [Karlsen et al., 2018] Karlsen, E., Schulz, C., and Almaas, E. (2018). Automated generation of genome-scale metabolic draft reconstruc-
1163 tions based on KEGG. *BMC Bioinformatics*, 19(1):467.
- 1164 [Karp et al., 2019a] Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker,
1165 M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M., and Subhraveti, P. (2019a). The BioCyc collection of microbial genomes and
1166 metabolic pathways. *Briefings in Bioinformatics*, 20(4):1085–1093.
- 1167 [Karp et al., 2019b] Karp, P. D., Midford, P. E., Billington, R., Kothari, A., Krummenacker, M., Latendresse, M., Ong, W. K.,
1168 Subhraveti, P., Caspi, R., Fulcher, C., Keseler, I. M., and Paley, S. M. (2019b). Pathway Tools version 23.0 update: Software
1169 for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*.
- 1170 [Karp et al., 2018] Karp, P. D., Weaver, D., and Latendresse, M. (2018). How accurate is automated gap filling of metabolic models?
1171 *BMC Systems Biology*, 12(1):73.
- 1172 [Kashaf et al., 2022] Kashaf, S. S., Proctor, D. M., Deming, C., Saary, P., Hölzer, M., Taylor, M. E., Kong, H. H., Segre, J. A., Almeida,
1173 A., and Finn, R. D. (2022). Integrating cultivation and metagenomics for a multi-kingdom view of skin microbiome diversity and
1174 functions. *Nature microbiology*, 7(1):169–179.
- 1175 [Kettle et al., 2018] Kettle, H., Holtrop, G., Louis, P., and Flint, H. J. (2018). microPop: Modelling microbial populations and commu-
1176 nities in R. *Methods in Ecology and Evolution*, 9(2):399–409.
- 1177 [Kim et al., 2022] Kim, C. Y., Ma, J., and Lee, I. (2022). HiFi metagenomic sequencing enables assembly of accurate and complete
1178 genomes from human gut microbiota. *Nature Communications*, 13(1):6367.
- 1179 [Kim et al., 2021] Kim, H.-M., Jeon, S., Chung, O., Jun, J. H., Kim, H.-S., Blazyte, A., Lee, H.-Y., Yu, Y., Cho, Y. S., Bolser, D. M., and
1180 Bhak, J. (2021). Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina
1181 sequencing benchmark for whole-genome sequencing. *GigaScience*, 10(3):giab014.
- 1182 [Kindler et al., 2021] Kindler, G. S., Wong, H. L., Larkum, A. W. D., Johnson, M., MacLeod, F. I., and Burns, B. P. (2021). Genome-
1183 resolved metagenomics provides insights into the functional complexity of microbial mats in Blue Holes, Shark Bay. *FEMS Microbiology
1184 Ecology*.

- 1185 [King et al., 2016] King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., and Lewis,
1186 N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*,
1187 44(D1):D515–D522.
- 1188 [Knight et al., 2018] Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T.,
1189 McCall, L.-I., McDonald, D., Melnik, A. V., Morton, J. T., Navas, J., Quinn, R. A., Sanders, J. G., Swafford, A. D., Thompson,
1190 L. R., Tripathi, A., Xu, Z. Z., Zaneveld, J. R., Zhu, Q., Caporaso, J. G., and Dorrestein, P. C. (2018). Best practices for analysing
1191 microbiomes. *Nature Reviews. Microbiology*, 16(7):410–422.
- 1192 [Kolmogorov et al., 2020] Kolmogorov, M., Bickhart, D. M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S. B., Kuhn, K., Yuan, J.,
1193 Polevikov, E., Smith, T. P. L., and Pevzner, P. A. (2020). metaFlye: Scalable long-read metagenome assembly using repeat graphs.
1194 *Nature Methods*, 17(11):1103–1110.
- 1195 [Koren et al., 2017] Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: Scalable and
1196 accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, 27(5):722–736.
- 1197 [Krakau et al., 2022] Krakau, S., Straub, D., Gourel, H., Gabernet, G., and Nahsen, S. (2022). Nf-core/mag: A best-practice pipeline
1198 for metagenome hybrid assembly and binning. *NAR Genomics and Bioinformatics*, 4(1):lqac007–.
- 1199 [Kreimer et al., 2012] Kreimer, A., Doron-Faigenboim, A., Borenstein, E., and Freilich, S. (2012). NetCmpt: A network-based tool for
1200 calculating the metabolic competition between bacterial species. *Bioinformatics (Oxford, England)*, 28(16):2195–2197.
- 1201 [Kumar et al., 2018] Kumar, M., Ji, B., Babaei, P., Das, P., Lappa, D., Ramakrishnan, G., Fox, T. E., Haque, R., Petri, W. A., Bäckhed,
1202 F., and Nielsen, J. (2018). Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: Lessons
1203 from genome-scale metabolic modeling. *Metabolic Engineering*, 49:128–142.
- 1204 [Kumar et al., 2019] Kumar, M., Ji, B., Zengler, K., and Nielsen, J. (2019). Modelling approaches for studying the microbiome. *Nature*
1205 *Microbiology*, 4(8):1253–1267.
- 1206 [Labarthe et al., 2019] Labarthe, S., Polizzi, B., Phan, T., Goudon, T., Ribot, M., and Laroche, B. (2019). A mathematical model to
1207 investigate the key drivers of the biogeography of the colon microbiota. *Journal of Theoretical Biology*, 462:552–581.
- 1208 [Lam et al., 2020] Lam, T. J., Stamboulian, M., Han, W., and Ye, Y. (2020). Model-based and phylogenetically adjusted quantification
1209 of metabolic interaction between microbial species. *PLoS Computational Biology*, 16(10):e1007951.
- 1210 [Langmead and Salzberg, 2012] Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*,
1211 9(4):357–359.
- 1212 [Lapidus and Korobeynikov, 2021] Lapidus, A. L. and Korobeynikov, A. I. (2021). Metagenomic Data Assembly – The Way of Decoding
1213 Unknown Microorganisms. *Frontiers in Microbiology*, 12.
- 1214 [Latendresse and Karp, 2018] Latendresse, M. and Karp, P. D. (2018). Evaluation of reaction gap-filling accuracy by randomization.
1215 *BMC Bioinformatics*, 19(1):53.
- 1216 [Leimbach et al., 2013] Leimbach, A., Hacker, J., and Dobrindt, U. (2013). E. coli as an all-rounder: The thin line between commensalism
1217 and pathogenicity. *Current Topics in Microbiology and Immunology*, 358:3–32.
- 1218 [Leipzig et al., 2021] Leipzig, J., Nüst, D., Hoyt, C. T., Ram, K., and Greenberg, J. (2021). The role of metadata in reproducible
1219 computational research. *Patterns*, 2(9):100322.

- 1220 [León and Nogales, 2022] León, D. S. and Nogales, J. (2022). Toward merging bottom-up and top-down model-based designing of
1221 synthetic microbial communities. *Current Opinion in Microbiology*, 69:102169.
- 1222 [Leslie Grady Jr et al., 1986] Leslie Grady Jr, CP., Gujer, W., Henze, M., Marais, G. v. R., and Tomonori, M. (1986). A model for
1223 single-sludge wastewater treatment systems. *Water science and technology*, 18(6):47–61.
- 1224 [Levy and Borenstein, 2012] Levy, R. and Borenstein, E. (2012). Advances in experimental medicine and biology. *Advances in experi-*
1225 *mental medicine and biology*, 751:329–345.
- 1226 [Levy and Borenstein, 2013] Levy, R. and Borenstein, E. (2013). Metabolic modeling of species interaction in the human microbiome
1227 elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences*, 110(31):12804–12809.
- 1228 [Levy et al., 2015] Levy, R., Carr, R., Kreimer, A., Freilich, S., and Borenstein, E. (2015). NetCooperate: A network-based tool for
1229 inferring host-microbe and microbe-microbe cooperation. *BMC Bioinformatics*, 16(1):1–6.
- 1230 [Lewis et al., 2010] Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm,
1231 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
1232 evolved *E. Coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*, 6(1):390.
- 1233 [Li et al., 2015] Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for
1234 large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England)*, 31(10):1674–1676.
- 1235 [Li, 2018] Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)*, 34(18):3094–3100.
- 1236 [Li and Durbin, 2009] Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioin-*
1237 *formatics (Oxford, England)*, 25(14):1754–1760.
- 1238 [Li et al., 2009] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup,
1239 . G. P. D. P. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16):2078–2079.
- 1240 [Lieven et al., 2020] Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., Bartell, J. A., Blank, L. M.,
1241 Chauhan, S., Correia, K., Diener, C., Dräger, A., Ebert, B. E., Edirisinghe, J. N., Faria, J. P., Feist, A. M., Fengos, G., Fleming, R.
1242 M. T., García-Jiménez, B., Hatzimanikatis, V., van Helvoirt, W., Henry, C. S., Hermjakob, H., Herrgård, M. J., Kaafarani, A., Kim,
1243 H. U., King, Z., Klamt, S., Klipp, E., Koehorst, J. J., König, M., Lakshmanan, M., Lee, D.-Y., Lee, S. Y., Lee, S., Lewis, N. E., Liu,
1244 F., Ma, H., Machado, D., Mahadevan, R., Maia, P., Mardinoglu, A., Medlock, G. L., Monk, J. M., Nielsen, J., Nielsen, L. K., Nogales,
1245 J., Nookaew, I., Palsson, B. O., Papin, J. A., Patil, K. R., Poolman, M., Price, N. D., Resendis-Antonio, O., Richelle, A., Rocha, I.,
1246 Sánchez, B. J., Schaap, P. J., Sheriff, R. S. M., Shoaie, S., Sonnenschein, N., Teusink, B., Vilaça, P., Vik, J. O., Wodke, J. A. H.,
1247 Xavier, J. C., Yuan, Q., Zakhartsev, M., and Zhang, C. (2020). MEMOTE for standardized genome-scale metabolic model testing.
1248 *Nature Biotechnology*, 38(3):272–276.
- 1249 [Lind and Pollard, 2021] Lind, A. L. and Pollard, K. S. (2021). Accurate and sensitive detection of microbial eukaryotes from whole
1250 metagenome shotgun sequencing. *Microbiome*, 9(1):58.
- 1251 [Liu et al., 2022] Liu, F., Heiner, M., and Gilbert, D. (2022). Hybrid modelling of biological systems: Current progress and future
1252 prospects. *Briefings in Bioinformatics*, 23(3):bbac081.
- 1253 [Lloréns-Rico et al., 2022] Lloréns-Rico, V., Simcock, J. A., Huys, G. R., and Raes, J. (2022). Single-cell approaches in human microbiome
1254 research. *Cell*, 185(15):2725–2738.

- 1255 [Lu et al., 2022] Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., Salzberg, S. L., and Steinegger, M.
1256 (2022). Metagenome analysis using the Kraken software suite. *Nature Protocols*, 17(12):2815–2839.
- 1257 [Lyalina et al., 2022] Lyalina, S., Stepanauskas, R., Wu, F., Sanjabi, S., and Pollard, K. S. (2022). Single cell genome sequencing
1258 of laboratory mouse microbiota improves taxonomic and functional resolution of this model microbial community. *PLoS ONE*,
1259 17(4):e0261795.
- 1260 [Machado et al., 2018] Machado, D., Andrejev, S., Tramontano, M., and Patil, K. R. (2018). Fast automated reconstruction of genome-
1261 scale metabolic models for microbial species and communities. *Nucleic Acids Research*, 46(15):7542–7553.
- 1262 [Machado et al., 2021] Machado, D., Maistrenko, O. M., Andrejev, S., Kim, Y., Bork, P., Patil, K. R., and Patil, K. R. (2021). Polarization
1263 of microbial communities between competitive and cooperative metabolism. *Nature Ecology & Evolution*, 5(2):195–203.
- 1264 [Mahadevan et al., 2002] Mahadevan, R., Edwards, J. S., and Doyle, F. J. (2002). Dynamic Flux Balance Analysis of Diauxic Growth
1265 in *Escherichia coli*. *Biophysical Journal*, 83(3):1331–1340.
- 1266 [Mahadevan and Schilling, 2003] Mahadevan, R. and Schilling, C. (2003). The effects of alternate optimal solutions in constraint-based
1267 genome-scale metabolic models. *Metabolic Engineering*, 5(4):264–276.
- 1268 [Maistrenko et al., 2020] Maistrenko, O. M., Mende, D. R., Luetge, M., Hildebrand, F., Schmidt, T. S. B., Li, S. S., Rodrigues, J. F. M.,
1269 von Mering, C., Coelho, L. P., Huerta-Cepas, J., Sunagawa, S., and Bork, P. (2020). Disentangling the impact of environmental and
1270 phylogenetic constraints on prokaryotic within-species diversity. *The ISME Journal*, 14(5):1247–1259.
- 1271 [Malik-Sheriff et al., 2019] Malik-Sheriff, R. S., Glont, M., Nguyen, T. V. N., Tiwari, K., Roberts, M. G., Xavier, A., Vu, M. T., Men, J.,
1272 Maire, M., Kananathan, S., Fairbanks, E. L., Meyer, J. P., Arankalle, C., Varusai, T. M., Knight-Schrijver, V., Li, L., Dueñas-Roca,
1273 C., Dass, G., Keating, S. M., Park, Y. M., Buso, N., Rodriguez, N., Hucka, M., and Hermjakob, H. (2019). BioModels—15 years of
1274 sharing computational models in life science. *Nucleic Acids Research*, 48(D1):D407–D415.
- 1275 [Mallick et al., 2019] Mallick, H., Franzosa, E. A., Mclver, L. J., Banerjee, S., Sirota-Madi, A., Kostic, A. D., Clish, C. B., Vlamakis, H.,
1276 Xavier, R. J., and Huttenhower, C. (2019). Predictive metabolomic profiling of microbial communities using amplicon or metagenomic
1277 sequences. *Nature Communications*, 10(1):3136.
- 1278 [Marsland et al., 2020] Marsland, R., Cui, W., and Mehta, P. (2020). A minimal model for microbial biodiversity can reproduce experi-
1279 mentally observed ecological patterns. *Scientific reports*, 10(1):1–17.
- 1280 [Massana and López-Escardó, 2022] Massana, R. and López-Escardó, D. (2022). Metagenome assembled genomes are for eukaryotes too.
1281 *Cell Genomics*, 2(5):100130.
- 1282 [Mendes-Soares et al., 2016] Mendes-Soares, H., Mundy, M., Soares, L. M., and Chia, N. (2016). MMinte: An application for predicting
1283 metabolic interactions among the microbial species in a community. *BMC Bioinformatics*, 17(1):1–10.
- 1284 [Mendoza et al., 2019] Mendoza, S. N., Olivier, B. G., Molenaar, D., and Teusink, B. (2019). A systematic assessment of current
1285 genome-scale metabolic reconstruction tools. *Genome Biology*, 20(1):158.
- 1286 [Menzel et al., 2016] Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with
1287 Kaiju. *Nature Communications*, 7(1):11257.
- 1288 [MetaHIT consortium et al., 2015] MetaHIT consortium, Forslund, K., Hildebrand, F., Nielsen, T., Falony, G., Chatelier, E. L., Suna-
1289 gawa, S., Prifti, E., Vieira-Silva, S., Gudmundsdottir, V., Pedersen, H. K., Arumugam, M., Kristiansen, K., Voigt, A. Y., Vestergaard,

- 1290 H., Hercog, R., Costea, P. I., Kultima, J. R., Li, J., Jørgensen, T., Levenez, F., Dore, J., Nielsen, H. B., Brunak, S., Raes, J., Hansen,
1291 T., Wang, J., Ehrlich, S. D., Bork, P., and Pedersen, O. (2015). Disentangling type 2 diabetes and metformin treatment signatures in
1292 the human gut microbiota. *Nature*, 528(7581):262–266.
- 1293 [Mikheenko et al., 2016] Mikheenko, A., Saveliev, V., and Gurevich, A. (2016). MetaQUAST: Evaluation of metagenome assemblies.
1294 *Bioinformatics (Oxford, England)*, 32(7):1088–1090.
- 1295 [Mistry et al., 2020] Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E.,
1296 Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., and Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic
1297 Acids Research*, 49(D1):gkaa913–.
- 1298 [Mitchell et al., 2020] Mitchell, K., Brito, J. J., Mandric, I., Wu, Q., Knyazev, S., Chang, S., Martin, L. S., Karlsberg, A., Gerasimov, E.,
1299 Littman, R., Hill, B. L., Wu, N. C., Yang, H. T., Hsieh, K., Chen, L., Littman, E., Shabani, T., Enik, G., Yao, D., Sun, R., Schroeder,
1300 J., Eskin, E., Zelikovsky, A., Skums, P., Pop, M., and Mangul, S. (2020). Benchmarking of computational error-correction methods
1301 for next-generation sequencing data. *Genome Biology*, 21(1):71.
- 1302 [Moletta, 2015] Moletta, R. (2015). *La Méthanisation*. Tec et Doc.
- 1303 [Monk et al., 2014] Monk, J., Nogales, J., and Palsson, B. O. (2014). Optimizing genome-scale network reconstructions. *Nature biotech-
1304 nology*, 32(5):447–52.
- 1305 [Moorthy et al., 2015] Moorthy, A. S., Brooks, S. P. J., Kalmokoff, M., and Eberl, H. J. (2015). A Spatially Continuous Model of
1306 Carbohydrate Digestion and Transport Processes in the Colon. *PLOS ONE*, 10(12):e0145309.
- 1307 [Moretti et al., 2016] Moretti, S., Martin, O., Van Du Tran, T., Bridge, A., Morgat, A., and Pagni, M. (2016). MetaNetX/MNXref –
1308 reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Research*,
1309 44(D1):D523–D526.
- 1310 [Muñoz-Tamayo et al., 2016] Muñoz-Tamayo, R., Giger-Reverdin, S., and Sauvant, D. (2016). Mechanistic modelling of in vitro fermen-
1311 tation and methane production by rumen microbiota. *Animal Feed Science and Technology*, 220:1–21.
- 1312 [Muñoz-Tamayo et al., 2010] Muñoz-Tamayo, R., Laroche, B., Walter, É., Doré, J., and Leclerc, M. (2010). Mathematical modelling of
1313 carbohydrate degradation by human colonic microbiota. *Journal of Theoretical Biology*, 266(1):189–201.
- 1314 [Narayan et al., 2020] Narayan, N. R., Weinmaier, T., Laserna-Mendieta, E. J., Claesson, M. J., Shanahan, F., Dabbagh, K., Iwai, S.,
1315 and DeSantis, T. Z. (2020). Piphillin predicts metagenomic composition and dynamics from DADA2-corrected 16S rDNA sequences.
1316 *BMC Genomics*, 21(1):56.
- 1317 [Nielsen et al., 2014] Nielsen, H. B., Almeida, M., Juncker, A. S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D. R., Gautier, L.,
1318 Pedersen, A. G., Chatelier, E. L., Pelletier, E., Bonde, I., Nielsen, T., Manichanh, C., Arumugam, M., Batto, J.-M., dos Santos,
1319 M. B. Q., Blom, N., Borruel, N., Burgdorf, K. S., Boumezeur, F., Casellas, F., Doré, J., Dworzynski, P., Guarner, F., Hansen, T.,
1320 Hildebrand, F., Kaas, R. S., Kennedy, S., Kristiansen, K., Kultima, J. R., Léonard, P., Levenez, F., Lund, O., Moumen, B., Paslier,
1321 D. L., Pons, N., Pedersen, O., Prifti, E., Qin, J., Raes, J., Sørensen, S., Tap, J., Tims, S., Ussery, D. W., Yamada, T., Nielsen,
1322 H. B., Almeida, M., Juncker, A. S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D. R., Gautier, L., Pedersen, A. G., Chatelier,
1323 E. L., Pelletier, E., Bonde, I., Nielsen, T., Manichanh, C., Arumugam, M., Batto, J.-M., dos Santos, M. B. Q., Blom, N., Borruel,
1324 N., Burgdorf, K. S., Boumezeur, F., Casellas, F., Doré, J., Dworzynski, P., Guarner, F., Hansen, T., Hildebrand, F., Kaas, R. S.,
1325 Kennedy, S., Kristiansen, K., Kultima, J. R., Leonard, P., Levenez, F., Lund, O., Moumen, B., Paslier, D. L., Pons, N., Pedersen,
1326 O., Prifti, E., Qin, J., Raes, J., Sørensen, S., Tap, J., Tims, S., Ussery, D. W., Yamada, T., Renault, P., Sicheritz-Ponten, T., Bork,

- 1327 P., Wang, J., Brunak, S., Ehrlich, S. D., Jamet, A., Mérieux, A., Cultrone, A., Torrejon, A., Quinquis, B., Brechot, C., Delorme,
1328 C., M'Rini, C., de Vos, W. M., Maguin, E., Varela, E., Guedon, E., Gwen, F., Haimet, F., Artiguenave, F., Vandemeulebrouck, G.,
1329 Denariáz, G., Khaci, G., Blottière, H., Knol, J., Weissenbach, J., Vlieg, J. E. T. v. H., Torben, J., Parkhill, J., Turner, K., van de
1330 Guchte, M., Antolin, M., Rescigno, M., Kleerebezem, M., Derrien, M., Galleron, N., Sanchez, N., Grarup, N., Veiga, P., Oozeer, R.,
1331 Dervyn, R., Layec, S., Bruls, T., Winogradski, Y., G, Z. E., Renault, P., Sicheritz-Ponten, T., Bork, P., Wang, J., Brunak, S., and
1332 Ehrlich, S. D. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using
1333 reference genomes. *Nature Biotechnology*, 32(8):822–828.
- 1334 [Nissen et al., 2021] Nissen, J. N., Johansen, J., Allesøe, R. L., Sønderby, C. K., Armenteros, J. J. A., Grønbech, C. H., Jensen, L. J.,
1335 Nielsen, H. B., Petersen, T. N., Winther, O., and Rasmussen, S. (2021). Improved metagenome binning and assembly using deep
1336 variational autoencoders. *Nature Biotechnology*, 39(5):555–560.
- 1337 [Noronha et al., 2018] Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Daniélsdóttir, A. D., Krecke,
1338 M., Merten, D., Haraldsdóttir, H. S., Heinken, A., Heirendt, L., Magnúsdóttir, S., Ravcheev, D. A., Sahoo, S., Gawron, P., Friscioni,
1339 L., Garcia, B., Prendergast, M., Puente, A., Rodrigues, M., Roy, A., Rouquaya, M., Wiltgen, L., Žagare, A., John, E., Krueger,
1340 M., Kuperstein, I., Zinovyev, A., Schneider, R., Fleming, R. M. T., and Thiele, I. (2018). The Virtual Metabolic Human database:
1341 Integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research*, 47(D1):D614–D624.
- 1342 [Norsigian et al., 2019] Norsigian, C. J., Pusarla, N., McConn, J. L., Yurkovich, J. T., Dräger, A., Palsson, B. O., and King, Z. (2019).
1343 BiGG Models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Research*.
- 1344 [Nurk et al., 2017] Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: A new versatile metagenomic
1345 assembler. *Genome Research*, 27(5):824–834.
- 1346 [Ofaim et al., 2017] Ofaim, S., Ofek-Lalzar, M., Sela, N., Jinag, J., Kashi, Y., Minz, D., and Freilich, S. (2017). Analysis of microbial
1347 functions in the rhizosphere using a metabolic-network based framework for metagenomics interpretation. *Frontiers in Microbiology*,
1348 8:1606.
- 1349 [O’Leary et al., 2016] O’Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B.,
1350 Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E.,
1351 Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W.,
1352 Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O’Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D.,
1353 Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D.,
1354 Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D., and Pruitt, K. D. (2016). Reference
1355 sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*,
1356 44(D1):D733–D745.
- 1357 [Olm et al., 2019] Olm, M. R., West, P. T., Brooks, B., Firek, B. A., Baker, R., Morowitz, M. J., and Banfield, J. F. (2019). Genome-
1358 resolved metagenomics of eukaryotic populations during early colonization of premature infants and in hospital rooms. *Microbiome*,
1359 7(1):26.
- 1360 [Opatovsky et al., 2018] Opatovsky, I., Santos-Garcia, D., Ruan, Z., Lahav, T., Ofaim, S., Mouton, L., Barbe, V., Jiang, J., Zchori-Fein,
1361 E., and Freilich, S. (2018). Modeling trophic dependencies and exchanges among insects’ bacterial symbionts in a host-simulated
1362 environment. *BMC Genomics*, 19(1):402.
- 1363 [Orakov et al., 2021] Orakov, A., Fullam, A., Coelho, L. P., Khedkar, S., Szklarczyk, D., Mende, D. R., Schmidt, T. S. B., and Bork, P.
1364 (2021). GUNC: Detection of chimerism and contamination in prokaryotic genomes. *Genome Biology*, 22(1):178.

- 1365 [Orth et al., 2010] Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nature biotechnology*, 28(3):245–248.
- 1366 [Overbeek et al., 2014] Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello,
1367 B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., and Stevens, R. (2014). The SEED and the rapid annotation of microbial
1368 genomes using subsystems technology (RAST). *Nucleic Acids Research*, 42(Database issue):D206–D214.
- 1369 [Özkurt et al., 2022] Özkurt, E., Fritscher, J., Soranzo, N., Ng, D. Y. K., Davey, R. P., Bahram, M., and Hildebrand, F. (2022). LotuS2:
1370 An ultrafast and highly accurate tool for amplicon sequencing analysis. *Microbiome*, 10(1):176.
- 1371 [Pan et al., 2023] Pan, S., Zhao, X.-M., and Coelho, L. P. (2023). SemiBin2: self-supervised contrastive learning leads to better MAGs
1372 for short- and long-read sequencing. *Bioinformatics*, 39(Suppl 1):i21–i29.
- 1373 [Pan et al., 2022] Pan, S., Zhu, C., Zhao, X.-M., and Coelho, L. P. (2022). A deep siamese neural network improves metagenome-
1374 assembled genomes in microbiome datasets across different environments. *Nature Communications*, 13(1):2326.
- 1375 [Parks et al., 2020] Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J., and Hugenholtz, P. (2020). A complete
1376 domain-to-species taxonomy for Bacteria and Archaea. *Nature Biotechnology*, pages 1–8.
- 1377 [Parks et al., 2015] Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: Assessing the
1378 quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7):1043–1055.
- 1379 [Peixoto et al., 2022] Peixoto, R. S., Voolstra, C. R., Sweet, M., Duarte, C. M., Carvalho, S., Villela, H., Lunshof, J. E., Gram, L.,
1380 Woodhams, D. C., Walter, J., Roik, A., Hentschel, U., Thurber, R. V., Daisley, B., Ushijima, B., Daffonchio, D., Costa, R., Keller-
1381 Costa, T., Bowman, J. S., Rosado, A. S., Reid, G., Mason, C. E., Walke, J. B., Thomas, T., and Berg, G. (2022). Harnessing the
1382 microbiome to prevent global biodiversity loss. *Nature Microbiology*, pages 1–10.
- 1383 [Peng et al., 2012] Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). IDBA-UD: A de novo assembler for single-cell and
1384 metagenomic sequencing data with highly uneven depth. *Bioinformatics (Oxford, England)*, 28(11):1420–1428.
- 1385 [Pérez-Cobas et al., 2020] Pérez-Cobas, A. E., Gomez-Valero, L., and Buchrieser, C. (2020). Metagenomic approaches in microbial
1386 ecology: An update on whole-genome and marker gene sequencing analyses. *Microbial Genomics*, 6(8).
- 1387 [Pessi et al., 2022] Pessi, I. S., Viitamäki, S., Virkkala, A.-M., Eronen-Rasimus, E., Delmont, T. O., Marushchak, M. E., Luoto, M., and
1388 Hultman, J. (2022). In-depth characterization of denitrifier communities across different soil ecosystems in the tundra. *Environmental*
1389 *Microbiome*, 17(1):30.
- 1390 [Pizarro et al., 2007] Pizarro, F., Varela, C., Martabit, C., Bruno, C., Pérez-Correa, J. R., and Agosin, E. (2007). Coupling kinetic
1391 expressions and metabolic networks for predicting wine fermentations. *Biotechnology and Bioengineering*, 98(5):986–998.
- 1392 [Plaza Oñate et al., 2019] Plaza Oñate, F., Le Chatelier, E., Almeida, M., Cervino, A. C. L., Gauthier, F., Magoulès, F., Ehrlich, S. D.,
1393 and Pichaud, M. (2019). MSPminer: Abundance-based reconstitution of microbial pan-genomes from shotgun metagenomic data.
1394 *Bioinformatics (Oxford, England)*, 35(9):1544–1552.
- 1395 [Popp and Centler, 2020] Popp, D. and Centler, F. (2020). μ BialSim: Constraint-Based Dynamic Simulation of Complex Microbiomes.
1396 *Frontiers in Bioengineering and Biotechnology*, 8:574.
- 1397 [Prigent et al., 2018] Prigent, S., Nielsen, J. C., Frisvad, J. C., and Nielsen, J. (2018). Reconstruction of 24 *Penicillium* genome-scale
1398 metabolic models shows diversity based on their secondary metabolism. *Biotechnology and Bioengineering*, 115(10):2604–2612.

- 1399 [Pronk and Medema, 2022] Pronk, L. J. U. and Medema, M. H. (2022). Whokaryote: Distinguishing eukaryotic and prokaryotic contigs
1400 in metagenomes based on gene structure. *Microbial Genomics*, 8(5):mgen000823.
- 1401 [Qin et al., 2010] Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada,
1402 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.,
1403 Batto, J.-M., Hansen, T., Paslier, D. L., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K.,
1404 Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner,
1405 F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T.,
1406 Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariáz, G., Dervyn, R., Forte, M., Friss, C., van de Guchte, M., Guedon, E.,
1407 Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Roux, K. L., Leclerc, M., Maguin,
1408 E., Minardi, R. M., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., de Vos, W., Winogradsky, Y., Zoetendal,
1409 E., Bork, P., Ehrlich, S. D., and Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing.
1410 *Nature*, 464(7285):59–65.
- 1411 [Quince et al., 2017a] Quince, C., Delmont, T. O., Raguideau, S., Alneberg, J., Darling, A. E., Collins, G., and Eren, A. M. (2017a).
1412 DESMAN: A new tool for de novo extraction of strains from metagenomes. *Genome Biology*, 18(1):181.
- 1413 [Quince et al., 2021] Quince, C., Nurk, S., Raguideau, S., James, R., Soyer, O. S., Summers, J. K., Limasset, A., Eren, A. M., Chikhi,
1414 R., and Darling, A. E. (2021). STRONG: Metagenomics strain resolution on assembly graphs. *Genome Biology*, 22(1):214.
- 1415 [Quince et al., 2017b] Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017b). Shotgun metagenomics, from
1416 sampling to analysis. *Nature Biotechnology*, 35(9):833–844.
- 1417 [Ramalho et al., 2022] Ramalho, T. P., Chopin, G., Pérez-Carrascal, O. M., Tromas, N., and Verseux, C. (2022). Selection of anabaena
1418 sp. PCC 7938 as a cyanobacterium model for biological ISRU on mars. *Applied and Environmental Microbiology*, 88(15).
- 1419 [Rawlings et al., 2017] Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2017). The MEROPS
1420 database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database.
1421 *Nucleic Acids Research*, 46(Database issue):gkx1134–.
- 1422 [Remien et al., 2021] Remien, C. H., Eckwright, M. J., and Ridenhour, B. J. (2021). Structural identifiability of the generalized
1423 Lotka–Volterra model for microbiome studies. *Royal Society Open Science*, 8(7):201378.
- 1424 [Richardson et al., 2022] Richardson, L., Allen, B., Baldi, G., Beracochea, M., Bileschi, M. L., Burdett, T., Burgin, J., Caballero-Pérez,
1425 J., Cochrane, G., Colwell, L. J., Curtis, T., Escobar-Zepeda, A., Gurbich, T. A., Kale, V., Korobeynikov, A., Raj, S., Rogers, A. B.,
1426 Sakharova, E., Sanchez, S., Wilkinson, D. J., and Finn, R. D. (2022). MGnify: The microbiome sequence data analysis resource in
1427 2023. *Nucleic Acids Research*.
- 1428 [Ruscheweyh et al., 2022] Ruscheweyh, H.-J., Milanese, A., Paoli, L., Karcher, N., Clayssen, Q., Metzger, M. I., Wirbel, J., Bork, P.,
1429 Mende, D. R., Zeller, G., and Sunagawa, S. (2022). Reference genome-independent taxonomic profiling of microbiomes with mOTUs3.
1430 *bioRxiv : the preprint server for biology*, page 2021.04.20.440600.
- 1431 [Ruuskanen et al., 2021] Ruuskanen, M. O., Åberg, F., Männistö, V., Havulinna, A. S., Méric, G., Liu, Y., Loomba, R., Vázquez-Baeza,
1432 Y., Tripathi, A., Valsta, L. M., Inouye, M., Jousilahti, P., Salomaa, V., Jain, M., Knight, R., Lahti, L., and Niiranen, T. J. (2021).
1433 Links between gut microbiome composition and fatty liver disease in a large population sample. *Gut microbes*, 13(1):1–22.
- 1434 [Sabater et al., 2022] Sabater, C., Calvete-Torre, I., Ruiz, L., and Margolles, A. (2022). Arabinoxylan and pectin metabolism in crohn’s
1435 disease microbiota: An in silico study. *International journal of molecular sciences*, 23(13).

- 1436 [Saier et al., 2021] Saier, M. H., Reddy, V. S., Moreno-Hagelsieb, G., Hendargo, K. J., Zhang, Y., Iddamsetty, V., Lam, K. J. K., Tian,
1437 N., Russum, S., Wang, J., and Medrano-Soto, A. (2021). The transporter classification database (TCDB): 2021 update. *Nucleic Acids*
1438 *Research*, 49(D1):gkaa1004–.
- 1439 [Scaglia et al., 2009] Scaglia, G. J., Aballay, P. M., Mengual, C. A., Vallejo, M. D., and Ortiz, O. A. (2009). Improved phenomenological
1440 model for an isothermal winemaking fermentation. *Food Control*, 20(10):887–895.
- 1441 [Schmieder and Edwards, 2011] Schmieder, R. and Edwards, R. (2011). Fast identification and removal of sequence contamination from
1442 genomic and metagenomic datasets. *PLoS ONE*, 6(3):e17288.
- 1443 [Scholz et al., 2016] Scholz, M., Ward, D. V., Pasolli, E., Tolio, T., Zolfo, M., Asnicar, F., Truong, D. T., Tett, A., Morrow, A. L.,
1444 and Segata, N. (2016). Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nature Methods*,
1445 13(5):435–438.
- 1446 [Schubert et al., 2016] Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: Rapid adapter trimming, identification,
1447 and read merging. *BMC Research Notes*, 9(1):88.
- 1448 [Scott et al., 2023] Scott, W. T., Benito-Vaquerizo, S., Zimmermann, J., Bajić, D., Heinken, A., Suarez-Diez, M., and Schaap, P. J.
1449 (2023). A structured evaluation of genome-scale constraint-based modeling tools for microbial consortia. *PLOS Computational*
1450 *Biology*, 19(8):e1011363.
- 1451 [Seaver et al., 2020] Seaver, S. M. D., Liu, F., Zhang, Q., Jeffries, J., Faria, J. P., Edirisinghe, J. N., Mundy, M., Chia, N., Noor, E.,
1452 Beber, M. E., Best, A. A., DeJongh, M., Kimbrel, J. A., D’haeseleer, P., McCorkle, S. R., Bolton, J. R., Pearson, E., Canon, S.,
1453 Wood-Charlson, E. M., Cottingham, R. W., Arkin, A. P., and Henry, C. S. (2020). The ModelSEED Biochemistry Database for
1454 the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and
1455 microbes. *Nucleic Acids Research*, pages gkaa746–.
- 1456 [Seemann, 2014] Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, 30(14):2068–
1457 2069.
- 1458 [Segata, 2018] Segata, N. (2018). On the Road to Strain-Resolved Comparative Metagenomics. *mSystems*, 3(2):e00190–17.
- 1459 [Sereika et al., 2022] Sereika, M., Kirkegaard, R. H., Karst, S. M., Michaelsen, T. Y., Sørensen, E. A., Wollenberg, R. D., and Albertsen,
1460 M. (2022). Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures
1461 and metagenomes without short-read or reference polishing. *Nature Methods*, 19(7):823–826.
- 1462 [Shen et al., 2016] Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file
1463 manipulation. *PLoS ONE*, 11(10):e0163962.
- 1464 [Shoaie et al., 2015] Shoaie, S., Ghaffari, P., Kovatcheva-Datchary, P., Mardinoglu, A., Sen, P., Pujos-Guillot, E., de Wouters, T., Juste,
1465 C., Rizkalla, S., Chilloux, J., Hoyles, L., Nicholson, J. K., Dore, J., Dumas, M. E., Clement, K., Bäckhed, F., and Nielsen, J. (2015).
1466 Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome. *Cell Metabolism*, 22(2):320–331.
- 1467 [Shorten et al., 2007] Shorten, R., Wirth, F., Mason, O., Wulff, K., and King, C. (2007). Stability criteria for switched and hybrid
1468 systems. *SIAM Review*, 49(4):545–592.
- 1469 [Sieber et al., 2018a] Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., and Banfield, J. F. (2018a).
1470 Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, 3(7):836–843.

- 1471 [Sieber et al., 2018b] Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., and Banfield, J. F. (2018b).
1472 Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, 3(7):836–843.
- 1473 [Silva et al., 2021] Silva, K. D., Pons, N., Berland, M., Oñate, F. P., Almeida, M., and Peterlongo, P. (2021). StrainFLAIR: Strain-level
1474 profiling of metagenomic samples using variation graphs. *PeerJ*, 9:e11884.
- 1475 [Smith et al., 2021] Smith, N. W., Shorten, P. R., Altermann, E., Roy, N. C., and McNabb, W. C. (2021). Examination of hydrogen
1476 cross-feeders using a colonic microbiota model. *BMC Bioinformatics*, 22(1):3.
- 1477 [Solden et al., 2016] Solden, L., Lloyd, K., and Wrighton, K. (2016). The bright side of microbial dark matter: Lessons learned from the
1478 uncultivated majority. *Current Opinion in Microbiology*, 31:217–226.
- 1479 [Somerville et al., 2019] Somerville, V., Lutz, S., Schmid, M., Frei, D., Moser, A., Irmeler, S., Frey, J. E., and Ahrens, C. H. (2019).
1480 Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain
1481 diversity and an active phage system. *BMC Microbiology*, 19(1):143.
- 1482 [Stanke et al., 2004] Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in
1483 eukaryotes. *Nucleic Acids Research*, 32(suppl_2):W309–W312.
- 1484 [Stein et al., 2013] Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Räscht, G., Pamer, E. G., Sander, C., and Xavier, J. B. (2013).
1485 Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota. *PLoS computational biology*,
1486 9(12):e1003388.
- 1487 [Steinegger and Söding, 2017] Steinegger, M. and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis
1488 of massive data sets. *Nature biotechnology*, 35(11):1026–1028.
- 1489 [Stelling et al., 2002] Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E. D. (2002). Metabolic network structure
1490 determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193.
- 1491 [Taffs et al., 2009] Taffs, R., Aston, J. E., Brileya, K., Jay, Z., Klatt, C. G., McGlynn, S., Mallette, N., Montross, S., Gerlach, R.,
1492 Inskip, W. P., Ward, D. M., and Carlson, R. P. (2009). In silico approaches to study mass and energy flows in microbial consortia:
1493 A syntrophic case study. *BMC Systems Biology*, 3(1):114.
- 1494 [Tamames and Puente-Sánchez, 2018] Tamames, J. and Puente-Sánchez, F. (2018). SqueezeMeta, A Highly Portable, Fully Automatic
1495 Metagenomic Analysis Pipeline. *Frontiers in Microbiology*, 9:3349.
- 1496 [Tchobanoglus et al., 2003] Tchobanoglus, G., Burton, F., and Stensel, H. D. (2003). Wastewater engineering: Treatment and reuse.
1497 *American Water Works Association. Journal*, 95(5):201.
- 1498 [Tedesoo et al., 2014] Tedesoo, L., Bahram, M., Pöhlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., Ruiz, L. V., Vasco-Palacios,
1499 A. M., Thu, P. Q., Suija, A., Smith, M. E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K.,
1500 Põldmaa, K., Piepenbring, M., Phosri, C., Peterson, M., Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou, A. L., Nilsson,
1501 R. H., Morgado, L. N., Mayor, J., May, T. W., Majuakim, L., Lodge, D. J., Lee, S. S., Larsson, K.-H., Kohout, P., Hosaka, K.,
1502 Hiiesalu, I., Henkel, T. W., Harend, H., Guo, L.-d., Greslebin, A., Grelet, G., Geml, J., Gates, G., Dunstan, W., Dunk, C., Drenkhan,
1503 R., Dearnaley, J., Kesel, A. D., Dang, T., Chen, X., Buegger, F., Brearley, F. Q., Bonito, G., Anslan, S., Abell, S., and Abarenkov, K.
1504 (2014). Global diversity and geography of soil fungi. *Science (New York, N. Y.)*, 346(6213):1256688.
- 1505 [Tekle et al., 2010] Tekle, Y. I., Grant, J. R., Kovner, A. M., Townsend, J. P., and Katz, L. A. (2010). Identification of new molecular
1506 markers for assembling the eukaryotic tree of life. *Molecular Phylogenetics and Evolution*, 55(3):1177–1182.

- 1507 [ten Hoopen et al., 2017] ten Hoopen, P., Finn, R. D., Bongo, L. A., Corre, E., Fosso, B., Meyer, F., Mitchell, A., Pelletier, E., Pesole,
1508 G., Santamaria, M., Willassen, N. P., and Cochrane, G. (2017). The metagenomic data life-cycle: Standards and best practices.
1509 *GigaScience*, 6(8):1–11.
- 1510 [Thiele and Palsson, 2010] Thiele, I. and Palsson, B. O. (2010). A protocol for generating a high-quality genome-scale metabolic recon-
1511 struction. *Nature Protocols*, 5(1):93–121.
- 1512 [Tito et al., 2019] Tito, R. Y., Chaffron, S., Caenepeel, C., Lima-Mendez, G., Wang, J., Vieira-Silva, S., Falony, G., Hildebrand, F., Darzi,
1513 Y., Rymenans, L., Verspecht, C., Bork, P., Vermeire, S., Joossens, M., and Raes, J. (2019). Population-level analysis of *Blastocystis*
1514 subtype prevalence and variation in the human gut microbiota. *Gut*, 68(7):1180.
- 1515 [Toya and Shimizu, 2013] Toya, Y. and Shimizu, H. (2013). Flux analysis and metabolomics for systematic metabolic engineering of
1516 microorganisms. *Biotechnology Advances*, 31(6):818–826.
- 1517 [Truong et al., 2017] Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C., and Segata, N. (2017). Microbial strain-level population
1518 structure and genetic diversity from metagenomes. *Genome Research*, 27(4):626–638.
- 1519 [Uritskiy et al., 2018] Uritskiy, G. V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved
1520 metagenomic data analysis. *Microbiome*, 6(1):158.
- 1521 [van den Berg et al., 2022] van den Berg, N. I., Machado, D., Santos, S., Rocha, I., Chacón, J., Harcombe, W., Mitri, S., and Patil, K. R.
1522 (2022). Ecological modelling approaches for predicting emergent properties in microbial communities. *Nature ecology & evolution*,
1523 6(7):855–865.
- 1524 [van Rosmalen et al., 2021] van Rosmalen, R., Smith, R., Martins dos Santos, V., Fleck, C., and Suarez-Diez, M. (2021). Model reduction
1525 of genome-scale metabolic models as a basis for targeted kinetic models. *Metabolic Engineering*, 64:74–84.
- 1526 [Varma and Palsson, 1994] Varma, A. and Palsson, B. O. (1994). Stoichiometric flux balance models quantitatively predict growth and
1527 metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, 60(10):3724–3731.
- 1528 [Vicedomini et al., 2021] Vicedomini, R., Quince, C., Darling, A. E., and Chikhi, R. (2021). Strainberry: Automated strain separation
1529 in low-complexity metagenomes using long reads. *Nature Communications*, 12(1):4485.
- 1530 [Wang and Jia, 2016] Wang, J. and Jia, H. (2016). Metagenome-wide association studies: Fine-mining the microbiome. *Nature Reviews*
1531 *Microbiology*, 14(8):508–522.
- 1532 [Watson and Warr, 2019] Watson, M. and Warr, A. (2019). Errors in long-read assemblies can critically affect protein prediction. *Nature*
1533 *Biotechnology*, 37(2):124–126.
- 1534 [Weiss et al., 2022] Weiss, A. S., Burrichter, A. G., Durai Raj, A. C., von Stempel, A., Meng, C., Kleigrewe, K., Münch, P. C., Rössler,
1535 L., Huber, C., Eisenreich, W., Jochum, L. M., Göing, S., Jung, K., Lincetto, C., Hübner, J., Marinos, G., Zimmermann, J., Kaleta, C.,
1536 Sanchez, A., and Stecher, B. (2022). In vitro interaction network of a synthetic gut bacterial community. *ISME Journal*, 16(4):1095–
1537 1109.
- 1538 [Wemheuer et al., 2020] Wemheuer, F., Taylor, J. A., Daniel, R., Johnston, E., Meinicke, P., Thomas, T., and Wemheuer, B. (2020).
1539 Tax4Fun2: Prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. *Environ-*
1540 *mental Microbiome*, 15(1):11.
- 1541 [West et al., 2007] West, S. A., Diggle, S. P., Buckling, A., Gardner, A., and Griffin, A. S. (2007). The social lives of microbes. *Annual*
1542 *Review of Ecology, Evolution, and Systematics*, 38(1):53–77.

- 1543 [Widmer and Stelling, 2018] Widmer, L. A. and Stelling, J. (2018). Bridging intracellular scales by mechanistic computational models.
1544 *Current Opinion in Biotechnology*, 52:17–24.
- 1545 [Wilkinson et al., 2016] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten,
1546 J.-W., Santos, L. B. d. S., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S.,
1547 Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., Hoen, P. A., Hooft, R.,
1548 Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik,
1549 R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen,
1550 E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The FAIR Guiding Principles for
1551 scientific data management and stewardship. *Scientific Data*, 3(1):160018.
- 1552 [Wilmes et al., 2022] Wilmes, P., Martin-Gallausiaux, C., Ostaszewski, M., Aho, V. T., Novikova, P. V., Laczny, C. C., and Schneider,
1553 J. G. (2022). The gut microbiome molecular complex in human health and disease. *Cell Host & Microbe*, 30(9):1201–1206.
- 1554 [Worden et al., 2015] Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., and Keeling, P. J. (2015).
1555 Environmental science. Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science (New York,*
1556 *N. Y.)*, 347(6223):1257594.
- 1557 [Wu et al., 2016] Wu, Y.-W., Simmons, B. A., and Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover
1558 genomes from multiple metagenomic datasets. *Bioinformatics (Oxford, England)*, 32(4):605–607.
- 1559 [Xavier et al., 2017] Xavier, J. C., Patil, K. R., and Rocha, I. (2017). Integration of biomass formulations of genome-scale metabolic
1560 models with experimental data reveals universally essential cofactors in prokaryotes. *Metabolic Engineering*, 39:200–208.
- 1561 [Ye et al., 2019] Ye, S. H., Siddle, K. J., Park, D. J., and Sabeti, P. C. (2019). Benchmarking Metagenomics Tools for Taxonomic
1562 Classification. *Cell*, 178(4):779–794.
- 1563 [Yilmaz et al., 2011] Yilmaz, P., Gilbert, J. A., Knight, R., Amaral-Zettler, L., Karsch-Mizrachi, I., Cochrane, G., Nakamura, Y.,
1564 Sansone, S.-A., Glöckner, F. O., and Field, D. (2011). The genomic standards consortium: Bringing standards to life for microbial
1565 ecology. *The ISME Journal*, 5(10):1565–1567.
- 1566 [Zampieri et al., 2023] Zampieri, G., Campanaro, S., Angione, C., and Treu, L. (2023). Metatranscriptomics-guided genome-scale
1567 metabolic modeling of microbial communities. *Cell Reports Methods*, page 100383.
- 1568 [Zaramela et al., 2021] Zaramela, L. S., Moyne, O., Kumar, M., Zuniga, C., Tibocha-Bonilla, J. D., and Zengler, K. (2021). The sum is
1569 greater than the parts: Exploiting microbial communities to achieve complex functions. *Current Opinion in Biotechnology*, 67:149–157.
- 1570 [Zelezniak et al., 2015] Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., and Patil, K. R. (2015). Metabolic dependen-
1571 cies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences*, 112(20):6449–6454.
- 1572 [Zengler and Zaramela, 2018] Zengler, K. and Zaramela, L. S. (2018). The social network of microorganisms — how auxotrophies shape
1573 complex communities. *Nature Reviews Microbiology*, 16(6):383–390.
- 1574 [Zhang et al., 2019] Zhang, X., Li, L., Butcher, J., Stintzi, A., and Figeys, D. (2019). Advancing functional and translational microbiome
1575 research using meta-omics approaches. *Microbiome*, 7(1):154.
- 1576 [Zheng et al., 2022] Zheng, W., Zhao, S., Yin, Y., Zhang, H., Needham, D. M., Evans, E. D., Dai, C. L., Lu, P. J., Alm, E. J., and Weitz,
1577 D. A. (2022). High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science (New*
1578 *York, N. Y.)*, 376(6597):eabm1483.

- 1579 [Zimmermann et al., 2021] Zimmermann, J., Kaleta, C., and Waschina, S. (2021). Gapseq: Informed prediction of bacterial metabolic
1580 pathways and reconstruction of accurate metabolic models. *Genome Biology*, 22(1):81.
- 1581 [Zomorodi and Maranas, 2012] Zomorodi, A. R. and Maranas, C. D. (2012). OptCom: A Multi-Level Optimization Framework for the
1582 Metabolic Modeling and Analysis of Microbial Communities. *PLoS Computational Biology*, 8(2):e1002363.
- 1583 [Zorrilla et al., 2021] Zorrilla, F., Buric, F., Patil, K. R., and Zelezniak, A. (2021). metaGEM: Reconstruction of genome scale metabolic
1584 models directly from metagenomes. *Nucleic Acids Research*, 49(21):e126–e126.