Establishing a Baseline for a Healthy Gut Microbiome

Artur Gomes[1] and Oscar Dias[2]

[1] University of Minho, Portugal
[2] University of Minho, Center of Biological Engineering (CEB), Portugal

Abstract. The human gut microbiome, a complex and dynamic ecosystem, plays a critical role in host health. Genome-scale metabolic modeling enables mechanistic exploration of microbial metabolism and interspecies interactions within this environment. In this study, we aim to integrate curated taxonomic profiles from GutFeelingKB with genome-scale metabolic models from the BiGG database to simulate the metabolic behavior of health-associated gut microbes. Using SMETANA, MICOM, and SteadyCom, we will explore metabolic cooperation, resource competition, and community stability. The resulting metabolic profiles will be analyzed using dimensionality reduction and clustering techniques to identify microbial ecotypes and metabolic subnetworks characteristic of a healthy gut. The main goal of this work is to build a modular simulation pipeline that can be used as a foundation for future efforts to characterize the gut microbiome and explore its functional organization.

Keywords: Gut Microbiome Health, Genome-scale metabolic models, Dysbiosis, Machine Learning, Metagenomics, Microbial community modeling

1 Contextualization and Motivation

1.1 Gut Microbiome in Health and Disease

The gut microbiome is a diverse and metabolically active microbial community that contributes to essential physiological processes, including digestion, nutrient absorption, immune regulation, and the production of bioactive compounds [1,2]. It helps maintain homeostasis, protects against pathogens, and modulates inflammatory responses [3]. Its composition is influenced by various factors that modulate microbial diversity and function, such as genetics [4], diet [5], age [6], geography [6], and medications [7], all of which contribute to differences in microbiome resilience and susceptibility to microbial imbalance often associated with disease, more commonly known as dysbiosis [8]. Dysbiosis has been linked to a wide range of diseases, including inflammatory bowel diseases [9,10], metabolic disorders like obesity and type 2 diabetes[11,12], and neurological conditions such as depression and neurodegeneration [13,14].

Despite considerable inter-individual variability, studies identified a possible "healthy core" microbiome, comprising taxa and metabolic functions that are consistently associated with health across individuals and populations[15].

Given the microbiome's influence on host health [16], there has been a surge of interest in microbiome-based interventions aimed at restoring microbial balance and reducing disease risk. Strategies such as pre and probiotics [17], fecal microbiota transplantation [18], and dietary modifications have been explored to modulate microbial communities and improve clinical outcomes [19]. In parallel to these therapeutic approaches, advances in high-throughput sequencing and computational biology have driven a broader integration of microbiome research into precision medicine and systems biology. These developments have enabled the identification of microbiome-derived biomarkers and the modeling of complex host–microbe interactions[20,21]. Yet, the high degree of interindividual variability and the functional complexity of microbial ecosystems remain major challenges. Addressing these requires integrative analytical frameworks that combine multi-omics data, metabolic modeling, and Machine Learning (ML) to better characterize microbiome in health and disease [20].

1.2 Aims

This project explores a pipeline for characterizing intestinal microbial communities through community-level metabolic simulations. Starting from taxonomic profiles curated in the Gut Feeling Knowledge Base (GutFeelingKB), we retrieved Genome-scale Metabolic Models (GEMs) for representative microbial strains from the BiGG Models database. These models were used to simulate microbial interactions and metabolic behavior using Species METabolic interaction ANAlysis (SMETANA), Microbial Community Modeling (MICOM), and Steady-state Community modeling (SteadyCom). To support exploratory analysis, simulation outputs from one representative condition were analyzed using dimensionality reduction and clustering to investigate functional differentiation among strains. The overall goal is to contribute toward a simulation-based framework for investigating the metabolic organization of gut microbial communities.

2 State-of-the-Art

2.1 Metagenomics in Microbiome Research

Metagenomics has transformed microbiome research by enabling the cultivation-independent analysis of microbial communities directly from environmental or host-associated samples[22,23]. Through high-throughput shotgun sequencing, metagenomics provides detailed insights into the taxonomic composition and functional potential of complex microbial ecosystems, including those inhabiting the human gut. Unlike targeted approaches such as 16S ribosome RNA gene sequencing, shotgun metagenomics captures information at the species and even strain level, allowing for more precise profiling of microbiome structure, which is essential for studying health-associated microbial patterns and for linking specific taxa to metabolic functions relevant to host physiology.

While Metagenome-Assembled Genomes (MAGs) have expanded the known diversity of unculturable gut microbes, reference-based profiling remains essential

for comparative studies and clinical applications. In this context, GutFeelingKB [24] provides a manually curated reference collection of microbial organisms consistently identified in the healthy human gut. GutFeelingKB is derived from metagenomics samples spanning the National Institute of Health (NIH) Human Microbiome Project [25], the George Washington University cohort, and other publicly available datasets, with a pipeline that integrates taxonomic profiling using CensuScope [26] and alignment against the Filtered database, followed by extensive manual validation to ensure high confidence taxonomic assignments.

While metagenomics provides insights into taxonomic composition and genetic potential, it cannot directly resolve dynamic metabolic activity, nutrient exchange, or pathway usage. The presence of a gene does not guarantee its expression, limiting functional predictions. To address these gaps, researchers integrate metagenomics with GEMs and multi-omics approaches [20]. By linking taxonomic profiles to GEMs, community-level metabolism can be simulated, enabling predictions of metabolic outputs and the functional roles of microbes in health and disease. This systems-level perspective bridges microbial composition with biological function.

2.2 Metabolic Modeling of the Gut Microbiome

GEMs address these limitations by representing microbial metabolism as stoichiometric networks, allowing simulations of biochemical fluxes under defined conditions. Their construction follows the principles of Constraint-Based Reconstruction and Analysis (COBRA) [27], a widely adopted framework for modeling metabolic networks. These models rely on the pseudo-steady-state assumption, which posits that the concentration of internal metabolites remains constant over time. Under this assumption, the system is typically analyzed using Flux Balance Analysis (FBA) [28], a linear programming approach that estimates flux distributions that satisfy the mass balance constraints while optimizing a predefined objective function, such as biomass production:

$$\max_{\mathbf{v}} \; c^T \cdot \mathbf{v} \quad \text{subject to} \quad \mathbf{S} \cdot \mathbf{v} = 0, \quad \mathbf{v}_{\min} \leq \mathbf{v} \leq \mathbf{v}_{\max} \tag{1}$$

In this formulation (Eq. 1), $\mathbf{c}$ is the objective coefficient vector that defines the biological goal of the simulation, such as maximizing biomass production. The steady-state assumption is encoded by the constraint $\mathbf{S} \cdot \mathbf{v} = 0$, where $\mathbf{S}$ is the stoichiometric matrix and $\mathbf{v}$ is the flux vector representing the rate of each reaction. The bounds $\mathbf{v}_{\min} \leq \mathbf{v} \leq \mathbf{v}_{\max}$ impose physiological or environmental constraints on each reaction, such as thermodynamic directionality, enzyme capacities, or nutrient availability. Solving this linear programming problem yields a feasible steady-state flux distribution that satisfies all constraints while optimizing the defined objective, thereby enabling quantitative predictions of metabolic behavior under specific conditions.

Several tools have been developed to automate or semi-automate the reconstruction process using genome annotations and biochemical databases. Notable examples include AGORA [29] and AGORA2 [30], curated collections of gut

microbial models; ModelSEED [31] and merlin [32], which facilitate draft reconstructions from genomic data; and CarveMe [33], a fast, template based tool for generating GEMs at scale. Additionally, tools like gapSeq [34] support gapfilling and refinement of metabolic networks to ensure model functionality. While such pipelines enable the generation of new models, curated repositories such as the BiGG Models [35] database provide a standardized and quality-controlled collection of GEMs for a wide range of microbial species. These resources support the integration of metagenomics data with predictive metabolic modeling in microbiome studies.

Beyond individual species modeling, community-level metabolic modeling allows exploration of interactions within complex microbial ecosystems such as the gut microbiome[20,36]. These approaches simulate metabolite exchange, resource competition, and cooperation, which shape community structure and function and generally follow one of two frameworks: steady-state or dynamic. Steady-state models assume constant internal metabolite concentrations and employ constraint-based formulations such as FBA and its community extensions. In contrast, dynamic approaches use ordinary differential equations to model temporal changes in biomass and metabolite levels, enabling simulations of ecological succession. Additionally, alternative methods have been developed that rely on different modeling strategies, providing complementary insights into microbial community behavior. Each approach offers a distinct perspective, enriching our understanding of these complex ecosystems. This study focuses on three representative steady-state modeling tools - SteadyCom, MICOM, and SMETANA - each of which offers a distinct perspective on simulating microbial communities.

SteadyCom [37] is a constraint-based algorithm designed for modeling microbial communities under ecological steady-state conditions. It represents the community as a multicompartment system in which each organism has its own compartment, and all species interact via a shared extracellular compartment that defines the environmental medium. The key assumption in SteadyCom is that all species in a stable community grow at a uniform, time-averaged growth rate, which reflects ecological equilibrium. Unlike classical FBA, fluxes in SteadyCom are not normalized per unit biomass but are instead scaled by species abundance.

A central feature of the method is the coupling of biomass flux and species abundance: an organism can only carry flux if it has nonzero biomass and growth rate, realistically excluding inactive or extinct species. The framework solves a convex optimization problem to compute the maximum feasible community growth rate while satisfying species-specific and community-level mass balance constraints. It is particularly well suited for modeling long-term microbial coexistence in environments like the human gut or industrial bioreactors. SteadyCom is also computationally efficient, requiring a fixed number of linear programs independent of the number of speciess. This is an advantage over dynamic models such as cFBA.

MICOM [38] extends constraint-based community modeling by incorporating species-specific relative abundances into a multi-objective optimization framework. Like SteadyCom, it assumes a shared extracellular environment and individual taxon compartments but does not impose the assumption of equal growth rates. Instead, MICOM introduces a cooperative trade-off strategy that balances the growth of individual taxa with the global community objective.

MICOM's two stage optimization begins by maximizing the weighted sum of taxon specific growth rates, where the weights correspond to observed abundances. A second stage then adjusts fluxes to maintain near optimal growth while optimizing community-level functions. The framework allows for simulation of dietary conditions, prediction of taxon-specific growth rates, and evaluation of metabolic interactions under defined environmental constraints. However, it requires reliable abundance data as input, and low-abundance taxa may be excluded if coverage is insufficient. MICOM does not natively support flux variability analysis and focuses on steady-state growth predictions rather than dynamic behaviors.

SMETANA [39] introduces an alternative framework for modeling microbial communities that infers metabolic interactions without relying on biomass optimization or predefined growth rates. While it assumes steady-state mass balance, it diverges from FBA based approaches like SteadyCom and MICOM by estimating cross-feeding and metabolic dependencies through pairwise metabolite exchange simulations in minimal media.

The method constructs growth-supporting media for each organism and evaluates whether other species can complement missing metabolic functions via extracellular metabolite exchange. This generates directional interaction networks and species-specific support scores that reflect obligate or facultative dependencies. SMETANA does not require abundance or growth rate data and scales well to large communities, making it well suited for exploring metabolic complementarity in natural or synthetic consortia.

2.3 Machine Learning in Microbiome Analysis

Analyzing microbiome metabolic models poses unique computational challenges due to the high dimensionality and complexity of microbial community interactions. While supervised machine learning methods (such as random forests and support vector machines) have demonstrated value in phenotype classification [40], this study emphasizes unsupervised approaches for exploratory discovery of metabolic patterns.

Dimensionality reduction techniques are particularly effective for this purpose [41]. These techniques project high-dimensional flux distributions into low-dimensional spaces, enabling the visualization and interpretation of complex metabolic profiles. When coupled with clustering algorithms, they uncover intrinsic groupings that may correspond to functional ecotypes or transitions between health and disease states.

This unsupervised framework offers three key advantages in the context of metabolic modeling: (1) it operates without requiring labeled training data, making it well suited for exploratory analysis; (2) it reveals dominant patterns in community-level metabolic behavior that emerge from simulation data; and (3) it enables the stratification of microbial communities into functional ecotypes, supporting biologically meaningful subgrouping beyond taxonomic classification[20]. Crucially, the biological interpretability of these methods ensures that the identified patterns retain mechanistic relevance to host–microbiome interactions.

2.4 Pipeline Overview and Tool Integration

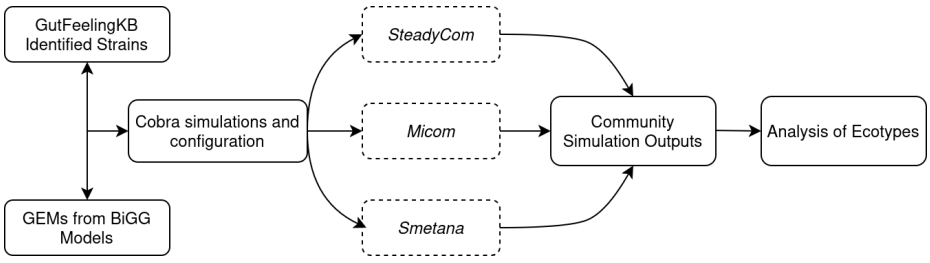

Fig. 1. Overview of the microbiome analysis pipeline.

Each tool contributes to a modular and extensible pipeline for exploring the metabolic structure of gut microbial communities (Figure 1). Taxonomic data from GutFeelingKB and metabolic models from the BiGG database were combined to generate a curated set of representative microbial strains and high-quality GEMs. These models were configured and simulated using SteadyCom, MICOM, and SMETANA, each offering complementary perspectives: SteadyCom estimated stable community composition under equal growth assumptions; MICOM simulated flux distributions using relative abundances; SMETANA identified potential metabolic dependencies between species without requiring abundance information. Outputs from these simulations were analyzed using techniques such as dimensionality reduction and clustering, with the goal of identifying patterns of metabolic behavior that may contribute to future definitions of microbial ecotypes. This workflow forms an initial step toward building a simulation-based framework for characterizing gut microbial communities.

3 Materials and Methods

This section outlines the computational pipeline developed and applied in this study. Each step was designed to ensure consistency and reproducibility. All analyses were implemented in Python (version 3.10.16).

3.1 Retrieval of Strain-Specific GEMs

In this step, we used data from the GutFeelingKB[24] and BiGG Models database[35] (version 1.6) to construct a functionally representative set of gut microbial strains. From GutFeelingKB, we compiled and filtered taxa from three related datasets: GFKB v4 master list, GFKB v4 abundance dataset, and GFKB v5 prediabetes list. While the v4 master list and v5 datasets primarily provided taxonomic information and cross-version presence/absence tracking, the v4 abundance dataset subset additionally included relative abundance values for a subset of species from the v4 master list. A summary of dataset dimensions, key metadata fields, and missing data is provided in Table 1.

Table 1. Summary of GutFeelingKB Datasets Used for Model Selection

Dataset	**Dimensions**	**Key Columns**	**Missing (%)**
GFKB v4 Master List	188 × 18	Unique ID, UP ID, presence flags, taxonomy	6.38
GFKB v4 Epilepsy Abundance	188 × 43	UP name, Lineage, Control/Case abundance statistics	1.98
GFKB v5 PreDiabetes	38 × 19	Unique ID, UP ID, presence flags, taxonomy	14.27

Data tables from GutFeelingKB were processed using the `pandas` Python library to standardize taxonomy and facilitate efficient cross-referencing between strain identifiers. Based on the strain identifiers present in the datasets, we performed cross-referencing with the BiGG Models database, which contains 108 manually curated GEMs serving as the reference set for model selection.

3.2 Model Validation and Minimal Medium Construction Using COBRApy

Each GEM retrieved from the BiGG database was evaluated using COBRApy[42] to confirm growth feasibility and ensure structural consistency. Biomass flux and metabolite exchange profiles were inspected via standard FBA simulations. For models that failed to grow under the default medium, the lower and upper bounds of relevant exchange reactions were manually adjusted to enable growth. Biomass reaction IDs were also standardized to support community modeling.

A minimal growth medium for community modeling was contructed by minimizing import fluxes for each strain while maintaining a biomass threshold. Two strategies were used: one starting from a fully open medium and another from the model's default configuration. The resulting strain-specific media were then merged by taking the sum import fluxes across all strains, yielding an aggregated-demand community medium.

3.3 Microbiome Model Construction and Simulation

To characterize microbial interactions within a representative gut community, we constructed community-scale metabolic simulations using three complementary modeling frameworks: SMETANA, MICOM, and SteadyCom.

SMETANA Simulations SMETANA[39] (version 1.2.1) was used to evaluate potential metabolic cooperation between community members. Each GEM was configured in individual SBML (XML) format, and simulations were run using both the global and detailed scoring algorithms under a complete medium. SMETANA computed pairwise species dependency scores and identified candidate cross-feeding metabolites, enabling the reconstruction of potential metabolic interactions. The resulting outputs were used to build a directed cross-feeding network, highlighting key metabolic donors and receivers within the community.

MICOM Simulations MICOM[38] (version 0.37.1) simulations were performed using a merged community model constructed from 9 of the 13 GEMs, corresponding to the strains for which abundance data was available in GutFeelingKB. Relative abundances derived from the GFKB v4 dataset were assigned to each strain. Baseline FBA was conducted using the cooperative trade-off strategy. From the output, we extracted individual strain growth rates and internal flux distributions, which were later used for dimensionality reduction and clustering analysis.

SteadyCom Simulations SteadyCom was used to estimate the steady-state composition of the microbial community under a defined medium condition obtained from the `cobrapy` step. All simulations were conducted using the `reframed` package (version 1.5.0).

4 Results and discussion

This section presents the main outcomes derived from each stage of the computational pipeline.

4.1 Selection of Representative GEMs

The taxonomic overlap across the three GutFeelingKB datasets, along with the subset of strains with GEMs available in the BiGG Models database, is illustrated in Figure 2. In the full dataset comparison (Figure 2.a), GFKB v4 and the epilepsy-specific subset share the majority of strains (160), with a few strains unique to each. GFKB v5 contributes 38 additional strains not found in the other sets. The Figure 2.b shows the filtered subset with GEM availability: most matched strains (9) originate from the GFKB v4 Epilepsy intersection, with a

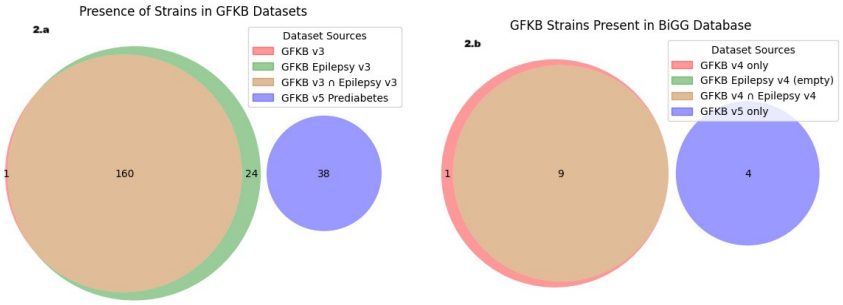

Fig. 2. (2a) Taxonomic overlap of strains across GutFeelingKB v4[24], the epilepsy-specific subset[43], and v5 (PreDiabetes)[44]. (2b) Subset of strains with GEMs available in the BiGG database. Most GEMs correspond to strains in the GFKB v4 ∩ Epilepsy intersection.

few unique to GFKB v4 (1) and GFKB v5 (4). No strains were uniquely available in BiGG from the epilepsy-specific dataset. This subset formed the basis for downstream metabolic modeling.

Out of the 14 candidate strains initially identified, 13 were retained for downstream simulation. One strain, the *Escherichia coli* IAI39, was excluded to better reflect the original distribution of taxa between the datasets. This decision was based on a proportional analysis aimed at preserving the dataset's composition (approximately 81% healthy, 19% prediabetic) while reducing the over representation of *E. coli*, a taxon already well covered by other selected strains. The final set of 13 curated GEMs is summarized in Table 2.

Table 2. Selected Strains with BiGG Models and GFKB Presence

UP Name	GFKB v4	GFKB v5	BiGG ID
E. coli BL21-DE3	Y	N	IEC1356_BL21DE3
E. coli SMS-3-5	Y	N	IECSMS35_1347
E. coli UT189	Y	N	UT189_1310
E. coli O17:K52:H18	Y	N	IECUMN_1333
E. coli O25b:H4-ST131	Y	N	IECSF_1327
E. coli O6H11	Y	N	IC_1306
E. coli O78H11	Y	N	IETEC_1333
E. coli OB3H11	Y	N	INR6857_1313
E. coli UMN688	Y	N	UMN688_1353
Peptoclostridium difficile	Y	N	KN900
E. coli SE11	N	Y	IECSF_1348
Klebsiella pneumoniae	N	Y	NY1228
Shigella dysenteriae s. 1	N	Y	ISDY_1059

Note: Y = Yes, N = No. Indicates presence in GFKB and BiGG model availability.

4.2 SMETANA Interaction Predictions

SMETANA simulation results revealed an asymmetric cross-feeding structure within the modeled community. Strain *iYL1228* acted as major metabolite donor. Figure 3 illustrates the potential distribution of exchanged metabolites and the potential role of specific strains in sustaining community-wide metabolic cooperation.

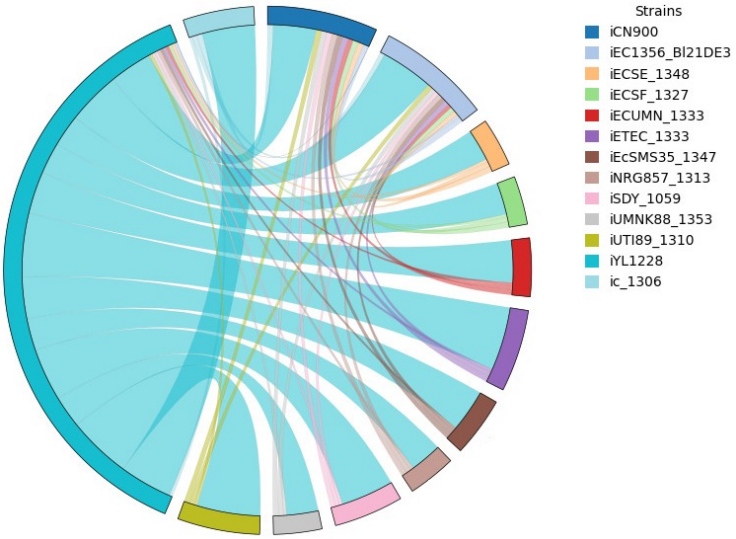

Fig. 3. Chord diagram representing pairwise metabolic dependencies predicted by SMETANA. Arcs indicate direction and magnitude of metabolite exchange between strains. Wider connections denote stronger dependencies. Donor strains such as *iYL1228* and *iETEC_ 1333* exhibit multiple outgoing links, supporting several recipient strains.

4.3 MICOM Community Behavior

To explore the metabolic differences among strains within the community, we performed a Principal Component Analysis (PCA) on the flux distribution matrix, followed by a 3D t-distributed Stochastic Neighbor Embedding (t-SNE) for nonlinear dimensionality reduction. The goal was to visualize strain-specific flux behavior in a reduced space.

Figure 4 shows the distribution of strains based on their predicted metabolic fluxes. Each point represents a strain and dispersion of points suggests that each strain exhibits a distinct metabolic flux profile. Notably, there is no apparent clustering, indicating that metabolic behaviors may be strain-specific under the given growth conditions.

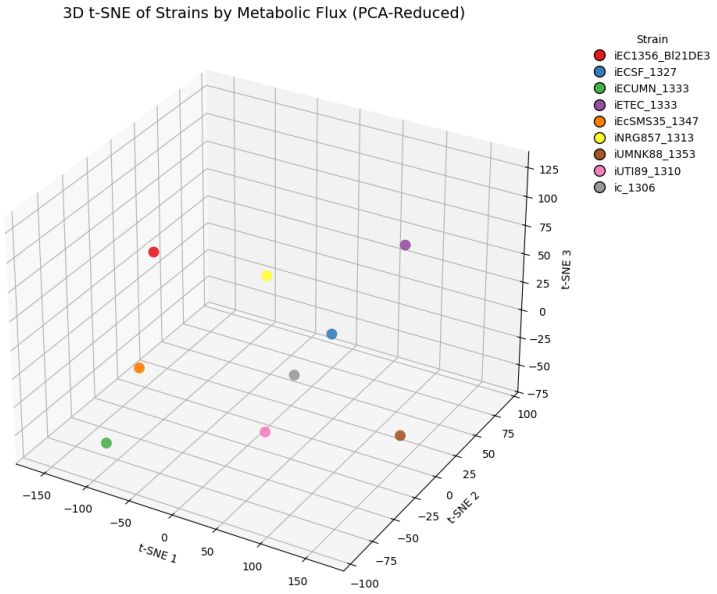

Fig. 4. 3D t-SNE plot of strains based on PCA-reduced flux distributions. Each point corresponds to a different strain. The spread of points suggests distinct metabolic behaviors without clear clustering.

4.4 SteadyCom-Predicted Community Composition

The abundances of individual strains were estimated using **SteadyCom** under a shared, defined *minimal medium*. This simulation approximates the contribution of each strain to a metabolically stable community under constrained nutrient availability.

As shown in Figure 5, most strains were assigned very low or zero abundance, indicating they do not significantly contribute to the stable community under the given nutrient conditions.

A few strains—notably **iYL1228** *(Klebsiella pneumoniae)*, **iETEC_1333** *(Escherichia coli 078:H11)*, and **iSDY_1059** *(Shigella dysenteriae)*—were assigned higher relative abundances in the SteadyCom simulation. This could indicate lower metabolic efficiency, as these strains require a larger biomass proportion to sustain the same community-wide growth rate as others. Alternatively, or additionally, their high abundance may reflect support roles. For instance, *Klebsiella pneumoniae* was identified in SMETANA results as a major donor strain, supporting multiple community members through metabolite exchange. Its elevated abundance may thus enable higher growth for the entire community by enhancing metabolic support to dependent strains.

Strain ID	Abundance
iYL1228	0.6238
iSDY_1059	0.1549
iCN900	0.0126
iECSE_1348	0.0
iUMNK88_1353	0.0022
iETEC_1333	0.1673
iECSF_1327	0.0
iEcSMS35_1347	0.0012
iEC1356_BL21DE3	0.0
iECUMN_1333	0.0365
iNRG857_1313	0.0013
iUTI89_1310	0.0
ic_1306	0.0001
Community growth	3.0254

Fig. 5. Abundances predicted by SteadyCom for each strain under a defined minimal medium. The values reflect each strain's contribution to the stable community.

5 Conclusion

This study served primarily as a proof-of-concept for integrating multiple community metabolic modeling tools into a reproducible pipeline. We successfully applied SMETANA, MICOM, and SteadyCom to a curated set of 13 strains and demonstrated that all frameworks can be effectively used to extract meaningful metabolic and ecological patterns. However, our analysis was constrained by the limited availability of curated GEMs for many gut-relevant taxa. The final dataset was heavily biased toward *Escherichia coli*, which, while useful for methodological testing, restricts the biological relevance. With that in mind, for future applications need to aim to expand taxonomic coverage, particularly, for key commensal and keystone species and try to simulate with more realistic gut like media.

References

1. Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., Gordon, J.I.: The human microbiome project 449(7164), 804–810, `https://www.nature.com/articles/nature06244`, publisher: Nature Publishing Group
2. Sender, R., Fuchs, S., Milo, R.: Revised estimates for the number of human and bacteria cells in the body 14(8), e1002533, `https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002533`, publisher: Public Library of Science

3. Thaiss, C.A., Zmora, N., Levy, M., Elinav, E.: The microbiome and innate immunity 535(7610), 65–74, https://www.nature.com/articles/nature18847, publisher: Nature Publishing Group
4. Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., Spector, T.D., Clark, A.G., Ley, R.E.: Human genetics shape the gut microbiome 159(4), 789–799, https://www.sciencedirect.com/science/article/pii/S0092867414012410
5. David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., Biddinger, S.B., Dutton, R.J., Turnbaugh, P.J.: Diet rapidly and reproducibly alters the human gut microbiome 505(7484), 559–563, https://www.nature.com/articles/nature12820, publisher: Nature Publishing Group
6. Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight, R., Gordon, J.I.: Human gut microbiome viewed across age and geography 486(7402), 222–227, https://www.nature.com/articles/nature11053, publisher: Nature Publishing Group
7. Dethlefsen, L., Relman, D.A.: Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation 108, 4554–4561, https://www.pnas.org/doi/10.1073/pnas.1000087107, publisher: Proceedings of the National Academy of Sciences
8. Carding, S., , Kristin, V., , Daniel T., V., , Bernard M., C., , Owen, L.J.: Dysbiosis of the gut microbiota in disease 26(1), 26191, https://www.tandfonline.com/doi/abs/10.3402/mehd.v26.26191, publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.3402/mehd.v26.26191
9. Frank, D.N., St. Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., Pace, N.R.: Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases 104(34), 13780–13785, https://www.pnas.org/doi/abs/10.1073/pnas.0706625104, publisher: Proceedings of the National Academy of Sciences
10. Ni, J., Wu, G.D., Albenberg, L., Tomov, V.T.: Gut microbiota and IBD: causation or correlation? 14(10), 573–584, https://www.nature.com/articles/nrgastro.2017.88, publisher: Nature Publishing Group
11. Ley, R.E., Turnbaugh, P.J., Klein, S., Gordon, J.I.: Human gut microbes associated with obesity 444(7122), 1022–1023, https://www.nature.com/articles/4441022a, publisher: Nature Publishing Group
12. Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., Peng, Y., Zhang, D., Jie, Z., Wu, W., Qin, Y., Xue, W., Li, J., Han, L., Lu, D., Wu, P., Dai, Y., Sun, X., Li, Z., Tang, A., Zhong, S., Li, X., Chen, W., Xu, R., Wang, M., Feng, Q., Gong, M., Yu, J., Zhang, Y., Zhang, M., Hansen, T., Sanchez, G., Raes, J., Falony, G., Okuda, S., Almeida, M., LeChatelier, E., Renault, P., Pons, N., Batto, J.M., Zhang, Z., Chen, H., Yang, R., Zheng, W., Li, S., Yang, H., Wang, J., Ehrlich, S.D., Nielsen, R., Pedersen, O., Kristiansen, K., Wang, J.: A metagenome-wide association study of gut microbiota in type 2 diabetes 490(7418), 55–60, https://www.nature.com/articles/nature11450, publisher: Nature Publishing Group
13. Cryan, J.F., Dinan, T.G.: Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour 13(10), 701–712, https://www.nature.com/articles/nrn3346, publisher: Nature Publishing Group

14. Sampson, T.R., Debelius, J.W., Thron, T., Janssen, S., Shastri, G.G., Ilhan, Z.E., Challis, C., Schretter, C.E., Rocha, S., Gradinaru, V., Chesselet, M.F., Keshavarzian, A., Shannon, K.M., Krajmalnik-Brown, R., Wittung-Stafshede, P., Knight, R., Mazmanian, S.K.: Gut microbiota regulate motor deficits and neuroinflammation in a model of parkinson's disease 167(6), 1469–1480.e12, https://www.cell.com/cell/abstract/S0092-8674(16)31590-2, publisher: Elsevier
15. Lloyd-Price, J., Abu-Ali, G., Huttenhower, C.: The healthy human microbiome 8(1), 51, https://doi.org/10.1186/s13073-016-0307-y
16. Lynch, S.V., Pedersen, O.: The human intestinal microbiome in health and disease 375(24), 2369–2379, https://www.nejm.org/doi/full/10.1056/NEJMra1600266, publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMra1600266
17. Gibson, G.R., Hutkins, R., Sanders, M.E., Prescott, S.L., Reimer, R.A., Salminen, S.J., Scott, K., Stanton, C., Swanson, K.S., Cani, P.D., Verbeke, K., Reid, G.: Expert consensus document: The international scientific association for probiotics and prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics 14(8), 491–502, https://www.nature.com/articles/nrgastro.2017.75, publisher: Nature Publishing Group
18. Wang, J.W., Kuo, C.H., Kuo, F.C., Wang, Y.K., Hsu, W.H., Yu, F.J., Hu, H.M., Hsu, P.I., Wang, J.Y., Wu, D.C.: Fecal microbiota transplantation: Review and update 118, S23–S31, https://www.sciencedirect.com/science/article/pii/S0929664618305552
19. Kovatcheva-Datchary, P., Nilsson, A., Akrami, R., Lee, Y.S., De Vadder, F., Arora, T., Hallen, A., Martens, E., Björck, I., Bäckhed, F.: Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of prevotella 22(6), 971–982, https://www.cell.com/cell-metabolism/abstract/S1550-4131(15)00517-3, publisher: Elsevier
20. Raajaraam, L., Raman, K.: Modeling microbial communities: Perspective and challenges 13(8), 2260–2270, https://doi.org/10.1021/acssynbio.4c00116, publisher: American Chemical Society
21. Heinken, A., Hulshof, T.O., Nap, B., Martinelli, F., Basile, A., O'Brolchain, A., O'Sullivan, N.F., Gallagher, C., Magee, E., McDonagh, F., Lalor, I., Bergin, M., Evans, P., Daly, R., Farrell, R., Delaney, R.M., Hill, S., McAuliffe, S.R., Kilgannon, T., Fleming, R.M., Thinnes, C.C., Thiele, I.: APOLLO: A genome-scale metabolic reconstruction resource of 247,092 diverse human microbes spanning multiple continents, age groups, and body sites p. 2023.10.02.560573, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10592896/
22. Zhou, Y., Liu, M., Yang, J.: Recovering metagenome-assembled genomes from shotgun metagenomic sequencing data: Methods, applications, challenges, and opportunities 260, 127023, https://www.sciencedirect.com/science/article/pii/S0944501322000635
23. Hanemaaijer, M., Röling, W.F.M., Olivier, B.G., Khandelwal, R.A., Teusink, B., Bruggeman, F.J.: Systems modeling approaches for microbial community studies: from metagenomics to inference of the community structure 6, https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2015.00213/full, publisher: Frontiers
24. King, C.H., Desai, H., Sylvetsky, A.C., LoTempio, J., Ayanyan, S., Carrie, J., Crandall, K.A., Fochtman, B.C., Gasparyan, L., Gulzar, N., Howell, P., Issa, N., Krampis, K., Mishra, L., Morizono, H., Pisegna, J.R., Rao, S., Ren, Y., Simonyan, V., Smith, K., VedBrat, S., Yao, M.D., Mazumder, R.: Baseline human

gut microbiota profile in healthy people and standard reporting template 14(9), e0206484, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0206484, publisher: Public Library of Science
25. Proctor, L.M., Creasy, H.H., Fettweis, J.M., Lloyd-Price, J., Mahurkar, A., Zhou, W., Buck, G.A., Snyder, M.P., Strauss, J.F., Weinstock, G.M., White, O., Huttenhower, C., The Integrative HMP (iHMP) Research Network Consortium: The integrative human microbiome project 569(7758), 641–648, https://www.nature.com/articles/s41586-019-1238-8, publisher: Nature Publishing Group
26. Shamsaddini, A., Pan, Y., Johnson, W.E., Krampis, K., Shcheglovitova, M., Simonyan, V., Zanne, A., Mazumder, R.: Census-based rapid and accurate metagenome taxonomic profiling 15(1), 918, https://doi.org/10.1186/1471-2164-15-918
27. Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., Magnusdóttir, S., Ng, C.Y., Preciat, G., Žagare, A., Chan, S.H.J., Aurich, M.K., Clancy, C.M., Modamio, J., Sauls, J.T., Noronha, A., Bordbar, A., Cousins, B., El Assal, D.C., Valcarcel, L.V., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Ben Guebila, M., Kostromins, A., Sompairac, N., Le, H.M., Ma, D., Sun, Y., Wang, L., Yurkovich, J.T., Oliveira, M.A.P., Vuong, P.T., El Assal, L.P., Kuperstein, I., Zinovyev, A., Hinton, H.S., Bryant, W.A., Aragón Artacho, F.J., Planes, F.J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M.A., Maranas, C.D., Lewis, N.E., Sauter, T., Palsson, B.O., Thiele, I., Fleming, R.M.T.: Creation and analysis of biochemical constraint-based models using the COBRA toolbox v.3.0 14(3), 639–702, https://www.nature.com/articles/s41596-018-0098-2, publisher: Nature Publishing Group
28. Orth, J.D., Thiele, I., Palsson, B.O.: What is flux balance analysis? 28(3), 245–248, https://www.nature.com/articles/nbt.1614, publisher: Nature Publishing Group
29. Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D.A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., Wilmes, P., Fleming, R.M.T., Thiele, I.: Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota 35(1), 81–89, https://www.nature.com/articles/nbt.3703, publisher: Nature Publishing Group
30. Heinken, A., Hertel, J., Acharya, G., Ravcheev, D.A., Nyga, M., Okpala, O.E., Hogan, M., Magnúsdóttir, S., Martinelli, F., Nap, B., Preciat, G., Edirisinghe, J.N., Henry, C.S., Fleming, R.M.T., Thiele, I.: Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine 41(9), 1320–1331, https://www.nature.com/articles/s41587-022-01628-0, publisher: Nature Publishing Group
31. Seaver, S.M.D., Liu, F., Zhang, Q., Jeffryes, J., Faria, J.P., Edirisinghe, J.N., Mundy, M., Chia, N., Noor, E., Beber, M.E., Best, A.A., DeJongh, M., Kimbrel, J.A., D'haeseleer, P., McCorkle, S.R., Bolton, J.R., Pearson, E., Canon, S., Wood-Charlson, E.M., Cottingham, R.W., Arkin, A.P., Henry, C.S.: The ModelSEED biochemistry database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes 49, D1555, https://doi.org/10.1093/nar/gkaa1143
32. Capela, J., Lagoa, D., Rodrigues, R., Cunha, E., Cruz, F., Barbosa, A., Bastos, J., Lima, D., Ferreira, E.C., Rocha, M., Dias, O.: merlin, an improved framework for the reconstruction of high-quality genome-scale metabolic models 50(11), 6052–6066, https://doi.org/10.1093/nar/gkac459

33. Machado, D., Andrejev, S., Tramontano, M., Patil, K.R.: Fast automated reconstruction of genome-scale metabolic models for microbial species and communities 46(15), 7542–7553, https://doi.org/10.1093/nar/gky537
34. Zimmermann, J., Kaleta, C., Waschina, S.: gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models 22(1), 81, https://doi.org/10.1186/s13059-021-02295-1
35. King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., Lewis, N.E.: BiGG models: A platform for integrating, standardizing and sharing genome-scale models 44, D515–D522, https://doi.org/10.1093/nar/gkv1049
36. Ibrahim, M., Raajaraam, L., Raman, K.: Modelling microbial communities: Harnessing consortia for biotechnological applications 19, 3892–3907, https://www.sciencedirect.com/science/article/pii/S2001037021002865
37. Chan, S.H.J., Simons, M.N., Maranas, C.D.: SteadyCom: Predicting microbial abundances while ensuring community stability 13(5), e1005539, https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005539, publisher: Public Library of Science
38. Diener, C., Gibbons, S.M., Resendis-Antonio, O.: MICOM: Metagenome-scale modeling to infer metabolic interactions in the gut microbiota 5(1), 10.1128/msystems.00606–19, https://journals.asm.org/doi/10.1128/msystems.00606-19, publisher: American Society for Microbiology
39. Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D.R., Bork, P., Patil, K.R.: Metabolic dependencies drive species co-occurrence in diverse microbial communities 112(20), 6449–6454, https://www.pnas.org/doi/full/10.1073/pnas.1421834112, publisher: Proceedings of the National Academy of Sciences
40. Hernández Medina, R., Kutuzova, S., Nielsen, K.N., Johansen, J., Hansen, L.H., Nielsen, M., Rasmussen, S.: Machine learning and deep learning applications in microbiome research 2(1), 98, https://doi.org/10.1038/s43705-022-00182-9
41. Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., Knight, R.: Bacterial community variation in human body habitats across space and time 326(5960), 1694–1697, https://www.science.org/doi/10.1126/science.1177486, publisher: American Association for the Advancement of Science
42. Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R.: COBRApy: COnstraints-based reconstruction and analysis for python 7(1), 74, https://doi.org/10.1186/1752-0509-7-74
43. Lindefeldt, M., Eng, A., Darban, H., Bjerkner, A., Zetterström, C.K., Allander, T., Andersson, B., Borenstein, E., Dahlin, M., Prast-Nielsen, S.: The ketogenic diet influences taxonomic and functional composition of the gut microbiota in children with severe epilepsy 5(1), 5, https://www.nature.com/articles/s41522-018-0073-2, publisher: Nature Publishing Group
44. Liu, Y., Wang, Y., Ni, Y., Cheung, C.K.Y., Lam, K.S.L., Wang, Y., Xia, Z., Ye, D., Guo, J., Tse, M.A., Panagiotou, G., Xu, A.: Gut microbiome fermentation determines the efficacy of exercise for diabetes prevention 31(1), 77–91.e5