

RESEARCH ARTICLE

Coarse-grained model of serial dilution dynamics in synthetic human gut microbiome

Tarun Mahajan ¹, Sergei Maslov ^{1,2,3,4*}

1 Department of Bioengineering, University of Illinois Urbana-Champaign, Urbana, Illinois, United States of America, **2** Center for Artificial Intelligence and Modeling at Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, Illinois, United States of America, **3** Department of Physics, University of Illinois Urbana-Champaign, Urbana, Illinois, United States of America, **4** NSF-Simons National Institute for Theory and Mathematics in Biology, Chicago IL

* maslov@illinois.edu OPEN ACCESS

Citation: Mahajan T, Maslov S (2025) Coarse-grained model of serial dilution dynamics in synthetic human gut microbiome. *PLoS Comput Biol* 21(7): e1013222. <https://doi.org/10.1371/journal.pcbi.1013222>

Editor: Nic Vega, Emory University Department of Biology, UNITED STATES OF AMERICA

Received: March 26, 2024

Accepted: June 10, 2025

Published: July 14, 2025

Copyright: © 2025 Mahajan, Maslov. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The consumption and production fluxes for the hCom2 strains used in this study were obtained from a previously published work: Han, S., Van Treuren, W., Fischer, C.R. et al., “A metabolomics pipeline for the mechanistic interrogation of the gut microbiome,” *Nature*, 595, 415–420 (2021). The dataset is available from the original publication at <https://doi.org/10.1038/s41586-021-03707-9>. The abundances of the hCom2 strains grown in mega media for multiple passages of a serial

Abstract

Many microbial communities in nature are complex, with hundreds of coexisting strains and the resources they consume. We currently lack the ability to assemble and manipulate such communities in a predictable manner in the lab. Here, we take a first step in this direction by introducing and studying a simplified consumer resource model of such complex communities in serial dilution experiments. The main assumption of our model is that during the growth phase of the cycle, strains share resources and produce metabolic byproducts in proportion to their average abundances and strain-specific consumption/production fluxes. We fit the model to describe serial dilution experiments in hCom2, a defined synthetic human gut microbiome with a steady-state diversity of 63 species growing on a rich media, using consumption and production fluxes inferred from metabolomics experiments. The model predicts serial dilution dynamics reasonably well, with a correlation coefficient between predicted and observed strain abundances as high as 0.8. We applied our model to: (i) calculate steady-state abundances of leave-one-out communities and use these results to infer the interaction network between strains; (ii) explore direct and indirect interactions between strains and resources by increasing concentrations of individual resources and monitoring changes in strain abundances; (iii) construct a resource supplementation protocol to maximally equalize steady-state strain abundances.

Author summary

Complex microbial communities, such as those in the human gut, are diverse ecosystems made up of hundreds of coexisting microbial strains that grow on a variety of nutrients. These communities often exist in environments characterized by “boom-and-bust”

dilution experiment were obtained from another previously published work: Jin, X., Yu, F.B., Yan, J. et al., "Culturing of a complex gut microbial community in mucin-hydrogel carriers reveals strain- and gene-associated spatial organization," *Nature Communications* 14, 3510 (2023). The dataset is available from the original publication at <https://doi.org/10.1038/s41467-023-39121-0>. Code used for analysis and visualization is available at https://github.com/Tarun-Mahajan/crn_gut_microbiome.

Funding: This research was supported in part by grants from the NSF (DMS-2235451) and Simons Foundation (MPS-NITMB-00005320) to the NSF-Simons National Institute for Theory and Mathematics in Biology (NITMB).

Competing interests: The authors have declared that no competing interests exist.

cycles, where nutrients are supplied in large batches before microbes undergo dilution or die-off. Traditional consumer-resource models struggle to capture the assembly dynamics of these complex communities, where interactions like resource competition and cross-feeding play a significant role. In our study, we addressed these challenges by developing a simplified consumer-resource model, which we tested on a synthetic human gut community (hCom2) containing 63 microbial species in serial dilution experiments. Using this model, we accurately predicted microbial population dynamics based on nutrient consumption and production data derived from metabolomics experiments. This approach enabled us to investigate how individual strains interact, assess the community's response to nutrient changes, and identify ways to balance species abundances by adjusting nutrient levels. Our model presents a valuable tool for understanding and potentially managing complex microbial communities.

Introduction

Many microbial communities in both natural [1–8] (human gut, plant rhizosphere, soil, ocean) and artificial or synthetic [9–14] settings are highly diverse, with the number of coexisting strains reaching into the hundreds. Such diversity of strains is often accompanied by the diversity of nutrients, with several hundred resources necessary to support the growth of hundreds of strains. In addition, many natural environments and most lab experiments on microbial communities are characterized by “boom-and-bust” cycles, where nutrients are added in large batches at the beginning of each cycle and species either die or get diluted in large ratios at the end of the cycle. Such serial dilution experiments are very easy to perform in the lab, but relatively difficult to model predictively [15–18].

There are a number of approaches suitable for modeling microbial communities with different levels of diversity. One of the most popular mathematical techniques uses generalized Lotka-Volterra (gLV) models [19–21], where strains are assumed to directly interact with each other. In gLV models, the exponential growth rate for each strain is approximated as a linear combination of the abundances of all strains in the community. However, this strategy does not explicitly account for resource competition or metabolic cross-feeding between strains. Thus, gLV models have been shown to be inadequate for the modeling of microbial communities with such indirect interactions [22].

Consumer-Resource Models (CRMs), such as the MacArthur model [23–25] or the Tilman model [26,27] are another popular choice for modeling microbial communities. These models explicitly describe shifts in community composition in response to changes in resource supply rates. CRMs also explicitly account for the consumption and production of metabolites by species. However, species and resources in the steady state of a classical CRM are assumed to be constant in time, which is appropriate for chemostat-like stable environments, but not for boom-and-bust dynamics.

The new generation of CRMs was developed to model microbial community dynamics in serial dilution and other strongly fluctuating environments [15–18,28]. CRMs with different levels of resolution are appropriate to describe communities with different levels of complexity. For example, we and others have developed the most detailed CRMs that can be used to describe low complexity communities with just a handful of strains and resources [15–17,29,30]. These models explicitly account for a variation in depletion times of individual resources and differences in growth rates and time lags of strains in each of the resulting temporal niches.

At the intermediate level of community complexity, Ho and collaborators recently proposed a CRM [28] for a simplified synthetic human gut microbiome consisting of 15 representative strains. Even at this reduced level of diversity, a number of approximations and simplifications were necessary. The authors clustered resources into 2^{15} groups based on the exact subset of species capable of consuming them, and then selected about 30 of the most abundant binary groups. However, this method is not scalable to the most complex communities with hundreds of species such as, e.g., hCom2 - a synthetic human gut microbiome with 120 strains developed in Ref [11].

Here, we present a simplified mechanistic model of complex microbial communities with hundreds of coexisting strains and resources. To test the performance of our model, we use it to predict the dynamics of 63 strains from hCom2 surviving in a serial dilution experiment [31]. The hCom2 community has been developed as a potential candidate for gut microbiome transplantation therapy [11]. It is therefore of practical importance to develop a reliable predictive model that can be used, for example, to predict the response of this synthetic community to various perturbations, or to attempt to equalize strain abundances prior to transplantation into the patient [32].

After fitting the parameters of the model based on the metabolomics data of Ref [33] and serial dilution data from Ref [31], we carried out three *in silico* perturbation experiments. First, we performed a leave-one-out experiment in which individual strains were removed from the inoculum one by one. This allowed us to resolve strain-strain interactions in the community as either cooperative or competitive. In the second experiment, we perturbed individual metabolites by increasing their concentrations in the medium and showed that this could both increase and decrease strain abundances. While an increase in the abundance of a strain in response to an increase in the concentration of a metabolite in the medium can be caused by both direct consumption of that metabolite and indirect effects such as cross-feeding on metabolic byproducts of other strains, a decrease in the abundance of a strain can only be caused by indirect effects such as competition for that and other metabolites with the rest of the strains in the community. Finally, based on the results of the second experiment, in the third experiment we proposed and implemented a greedy algorithm aimed at equalizing strain abundances in the community by increasing the concentrations of multiple metabolites.

Model and results

Consumer resource model of a complex community in serial dilution experiments

We introduce a simplified Consumer Resource Model (CRM) to predict the dynamics of a complex community composed of multiple microbial strains in the presence of multiple resources in serial dilution experiments. Consider n_S strains growing on n_R resources. The strains are grown in a serial dilution experiment, where the community culture is serially passaged. At the beginning of a passage, all strains are diluted by the same factor D . After that, the strains grow exponentially while consuming different resources. The abundance of a resource i is given by R_i , measured in units of its contribution to biomass. Thus we assume that yields of the same resource to the biomass of different strains are equal to each other and without loss of generality can be rescaled to 1. The existence of species-specific metabolic byproducts means that this assumption is only approximately true at best. Indeed, the production of a metabolic byproduct generally reduces the yield of the species growing on its precursor. However, in the absence of detailed information on which precursors were used to generate which byproducts, there is no way to correct for this effect. While the resource i is present in the environment, it is consumed by the strain α in proportion to α 's abundance

$N_\alpha(t)$ and strain-specific consumption flux, $c_{\alpha i}$ calculated per unit of its biomass. When multiple strains consume the same resource, the fraction consumed by the strain α is given by $\int_0^{T_i} c_{\alpha i} N_\alpha(t) / \int_0^{T_i} \sum_{\gamma=1}^{n_S} c_{\gamma i} N_\gamma(t)$. Here T_i is the time since the start of the growth cycle when the resource i is depleted. With these assumptions, mass conservation at the end of the growth cycle can be written as

$$N_\alpha(\text{end of cycle}) - N_\alpha(\text{start of cycle}) = \sum_{i=1}^{n_R} R_i \cdot \underbrace{\frac{c_{\alpha i} N_\alpha(\text{wt. avg. of cycle})}{\sum_{\gamma=1}^{n_S} c_{\gamma i} N_\gamma(\text{wt. avg. of cycle})}}_{\text{fraction of resource } i \text{ consumed by strain } \alpha} \tag{1}$$

Above we assumed that at any passage of a serial dilution experiment all resources get completely depleted. $N_\alpha(\text{start of cycle})$ and $N_\alpha(\text{end of cycle})$ are the abundances of strain α at the start and end of the growth cycle, respectively. $N_\alpha(\text{wt. avg. of cycle})$ is the weighted-average population size during the growth cycle, which determines the proportions for sharing resources among the strains. In principle, resources are consumed throughout the growth cycle, and an accurate calculation of $N_\alpha(\text{wt. avg. of cycle})$ requires integration of the instantaneous abundance over the entire growth cycle, but these data are usually not available. Assuming exponential growth with an approximately constant growth rate between 0 and T_i , we approximate $N_\alpha(\text{wt. avg. of cycle})$ by the geometric mean of $N_\alpha(\text{start of cycle})$ and $N_\alpha(\text{end of cycle})$:

$$N_\alpha(\text{wt. avg. of cycle}) \approx N_\alpha(\text{start of cycle})^{1-f} N_\alpha(\text{end of cycle})^f \tag{2}$$

Here a single parameter, $f \in [0, 1]$, which we call the time fraction, approximately accounts for two effects; (i) the fact that some nutrients get depleted prior to the end of the growth cycle, when the last nutrient gets depleted, (ii) the fact that $N_\alpha(\text{start of cycle}) < N_\alpha(\text{wt. avg. of cycle}) < N_\alpha(\text{end of cycle})$. For exponentially growing species, all resources are depleted towards the end of the growth cycle, thus we expect f to be close to 1 (see S1 Text). In principle, depletion times T_i differ between resources, so the time fraction f_i should also be resource specific. However, the relatively limited amount of training data available prevented us from adding 98 additional parameters for resource-specific f to our model. Therefore, we simplified the model to use the same time fraction f for all resources. Later, we demonstrate a simple method for estimating f from the experimental data.

In Eq (2), we assume that initial resource concentrations, R_i , in the medium are much higher than the Monod constants of any microbes consuming them. This condition is met for all resources in our analysis, as typical Monod constants range from micromolar to millimolar, which are too low to significantly contribute to a strain’s biomass. Under this assumption, a microbe consumes a resource at a rate independent of its concentration until the resource is nearly depleted. After a brief transient period — which we disregard in our calculations — the resource is fully exhausted, prompting species to switch to the next resource in their hierarchy.

At the beginning of a passage, after the end of the previous passage, all strain abundances are diluted by the factor D . Then, for any two consecutive passages, $k-1$ and k , $N_\gamma(\text{start of cycle } k) = N_\gamma(\text{end of cycle } k - 1)/D$. Now, substituting this and Eq (2) into Eq (1), we get Eq (3). There, we have also introduced interactions between strains through cross-feeding, and this is implemented as a multiplicative factor converting the concentration R_i of a resource i in the bolus medium to its total concentration ultimately consumed by the strains $R_i^{(\text{total})}$. The multiplier described by Eq (3b) depends on the normalized production flux, $p_{\gamma i}$,

calculated per unit of biomass of the strain γ producing the resource i . If a resource i is not produced as a metabolic byproduct by any strain, then $p_{\gamma i} = 0$, and $R_i^{(total)} = R_i$. However, if i is produced by at least one strain, then $p_{\gamma i} > 0$ for some γ , and $R_i^{(total)} > R_i$. If $D \gg 1$, then the second term on the left hand side (LHS) in Eq (3a) can be omitted. Eq (3) is the version of the model used for all analyses in the paper.

$$N_{\alpha}(\text{end of cycle } k) - \frac{N_{\alpha}(\text{end of cycle } k - 1)}{D} = \sum_{i=1}^{n_R} R_i^{(total)} \frac{c_{\alpha i} N_{\alpha}(\text{end of cycle } k - 1)^{1-f} N_{\alpha}(\text{end of cycle } k)^f}{\sum_{\gamma=1}^{n_S} c_{\gamma i} N_{\gamma}(\text{end of cycle } k - 1)^{1-f} N_{\gamma}(\text{end of cycle } k)^f}, \tag{3a}$$

$$R_i^{(total)} = R_i \left(1 + \underbrace{\sum_{\gamma=1}^{n_S} p_{\gamma i} N_{\gamma}(\text{end of cycle } k)}_{\text{cross-feeding}} \right) \tag{3b}$$

Our treatment of cross-feeding interactions is again a necessary simplification of what is generally a more complex dynamic process. We ignore possible differences in the timing of the production of metabolic by-products by using the biomass of the producing strains at the end of the cycle as a multiplier in Eq (3b). This approximation is justified in the case of rapid exponential growth where the average biomass is close to its value at the end of the cycle. Solving Eqs (3a) and (3b) for $N_{\alpha}(\text{end of cycle } k)$ as a function of $N_{\alpha}(\text{end of cycle } k - 1)$ defined passage-to-passage dynamics in serial dilution experiments. One can also set $N_{\alpha}(\text{end of cycle } k) = N_{\alpha}(\text{end of cycle } k - 1)$ and solve for steady-state abundances reached after multiple passages.

An illustration of the model is shown in Fig 1. In this example, there are two strains, A and B, and three resources, R_1 , R_2 , and R_3 . Both strains consume R_1 and produce R_2 (Fig 1a). However, neither strain consumes or produces R_3 (Fig 1a). R_1 is divided among the strains in proportion to their average abundances, $N_A(\text{avg.})$ and $N_B(\text{avg.})$, respectively, weighted by the respective consumption fluxes for R_1 , c_{A1} and c_{B1} (Fig 1b). The cross-feeding term involving R_2 is proportional to the abundances of strains A and B at the end of the growth cycle, $N_A(\text{end})$ and $N_B(\text{end})$, respectively. The contributions of A and B to the cross-feeding term are also weighted by the respective production fluxes for R_2 , p_{A2} and p_{B2} (Fig 1c).

Application of the model to predict serial dilution dynamics of a complex synthetic human gut microbial community

Data for fitting the consumer resource model. We have applied our model in Eq (3) to explain the serial dilution dynamics of a complex synthetic human gut microbial community [31] grown in a rich media (mega media) using two published and publicly available data sets [31,33]. Both datasets contain strains of a defined hCom2 community that was initially populated with prevalent bacterial strains from the human gut microbiome and subsequently challenged with a human fecal sample to fill open niches, resulting in increased stability to fecal challenge and robust colonization resistance (see Ref [11] for details). While the community in Ref [31]’s serial dilution experiments largely matches hCom2, it includes additional strains. Our model was applied to a subset of strains present in both this expanded list and hCom2. We refer to this subset community as hCom2* throughout the paper. One of the potential applications of such a community is gut microbiome transplantation or supplementation, which may have therapeutic implications for various diseases.

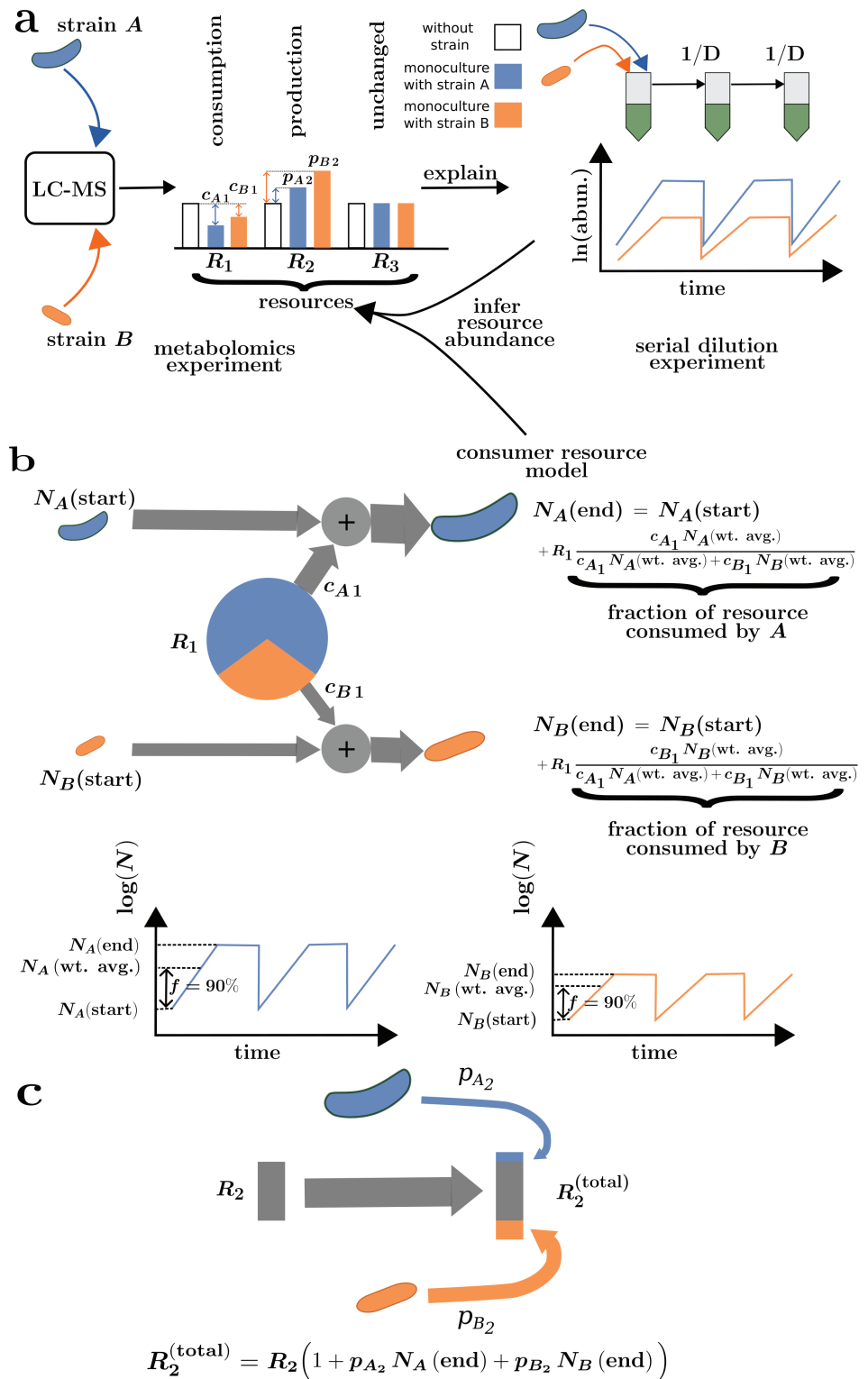


Fig 1. A schematic of our consumer resource model. This example has 2 strains (A and B) and 3 resources (R_1 , R_2 , and R_3) in a hypothetical community. **a**) (Left) Production and consumption fluxes are measured on a strain-by-strain basis from, e.g. a metabolomics experiment using a liquid chromatography-mass spectrometry

(LC-MS) technique. A unit of biomass of strains *A* and *B* consumes R_1 with consumption fluxes c_{A1} and c_{B1} , respectively, and produces R_2 as a byproduct with production fluxes p_{A2} and p_{B2} , respectively. Neither strain consumes or produces R_3 . (Right) Consumption and production fluxes are calculated as the difference between white and colored bars. Abundances for the strains in the community at multiple passages are obtained from a serial dilution experiment where abundances are diluted by a factor D at the beginning of the growth cycle. b) Rule for sharing of a resource between strains: In our model, we assume that during the growth cycle for each passage of the serial dilution experiment, R_1 is divided between *A* and *B* in proportion to their weighted average abundances N_A (wt. avg.) and N_B (wt. avg.), respectively, and to their consumption fluxes c_{A1} and c_{B1} , respectively. The weighted average strain abundances are obtained as the weighted geometric mean of the abundances at the beginning and end of the growth cycle as given by Eq (2). The weight for the geometric mean is called the time fraction f , and we estimated $f = 0.9$ or 90% of the growth cycle, which is shown in the log-abundance versus time plot here. c) Rule for production of a resource as byproduct by strains: In our model, we assume that for all strains in the community that consume R_2 , its excess concentration is contributed by its production as a byproduct by *A* and *B*. The contributions of *A* and *B* to this excess are proportional to their abundances at the end of the growth cycle N_A (end) and N_B (end), respectively, and to their normalized production fluxes p_{A2} and p_{B2} , respectively.

<https://doi.org/10.1371/journal.pcbi.1013222.g001>

Fitting our model to hCom2* involved estimating resource abundances R_i , which in this case represent different metabolites. To infer R_i , we collected consumption and production fluxes ($c_{\alpha i}$ and $p_{\gamma i}$, respectively) of 63 strains for 292 metabolites from the metabolomics experiments of Han et al. 2021 [33]. The names of the 63 strains along with the abbreviations we assigned are given in S1 Table. Details on the data set and the calculation of $c_{\alpha i}$ and $p_{\gamma i}$ are given in the Methods section.

In addition, we obtained the strain abundances N_α (end of cycle) for hCom2* grown in mega media for multiple passages of a serial dilution experiment from Jin et al. 2023 [31]. For model fitting, we used changes in strain abundances between the first and the second passage of the serial dilution experiment. This passage had the most dynamics of all the passages where we could separately study biological replicates [31] and was therefore the most suitable for fitting our dynamic model. The change from inoculum to first passage was less suitable because the inoculum data had only two biological replicates and these were not matched to three biological replicates in the first passage. We fitted the model using 189 data points (three biological replicates for the abundances of 63 strains) to fit Eq (3) for R_i . Details for this dataset are also given in the Methods section.

Coarse-graining the metabolite consumption and production fluxes. As discussed above, the number of available data points to fit the model is 189, which is less than the number of metabolites, 292. Thus, we coarse-grained the metabolites by clustering them by the consumption fluxes, which resulted in 98 clusters, including 10 non-singleton and 88 singleton clusters. The average consumption and production fluxes of the 63 strains were then calculated for these 98 metabolite clusters. The clustering process and the estimation of consumption and production fluxes for the metabolite clusters are described in detail in the Methods section. A clustered heatmap for the 98 metabolite clusters is shown in S2 Fig. The names of the metabolites in each metabolite cluster are given in S2 Table. In the rest of the manuscript, we use the terms metabolite and metabolite cluster synonymously, unless otherwise indicated.

Estimating the time fraction f . For the synthetic human gut microbial community hCom2*, we performed a systematic search for the time fraction f used in Eqs (2) and (3) and found that the predictive performance of our fitted model shows a peak around $f = 0.9$ (Methods, S1 Fig). A value close to 1 is expected for rapidly growing bacteria and experiments with large dilution factors ($D = 15,000$ in Ref [31]), where resources tend to be depleted towards the end of the growth cycle. Details of the estimation procedure are given in the Methods section.

Fitting and validating the model on the synthetic human gut microbiome. After obtaining the consumption and production fluxes, time fraction, and using abundances at passages 1 and 2 of the serial dilution experiment, we inferred R_i for the metabolite clusters by plugging these values into our model in Eq (3). Details of the fitting procedure are given in the Methods section. Of the 98 metabolite clusters, 38 had non-zero ($\geq 10^{-8}$) fitted R_i .

Using the fitted values of R_i , we validated the model through an independent in silico prediction experiment, which consisted of predicting the dynamics of strain abundances all the way from the inoculum to steady state reached around passage 3. To do this, we started with the experimentally measured strain abundances in the inoculum (average over two biological replicates in Ref [31]). Since the inoculum data were not used in our fitting procedure, starting from it helps to obtain an independent estimate of model performance. Using the inoculum abundances as N_α (end of cycle 0) Eq (3) with the inferred R_i was iteratively solved to obtain N_α (end of cycle 1), N_α (end of cycle 2), and N_α (end of cycle 3). The details of solving Eq (3) for prediction are described in the Methods section. We stopped the model at passage 3, since the experimental data suggest that the community reaches steady state by then [31].

A comparison between the predicted and observed abundances from the above in silico experiment is shown for passages 1 to 3 in Fig 2. The predicted and observed abundances are strongly correlated at passages 2 (Pearson's correlation coefficient = 0.77, $p = 1.58 \times 10^{-13}$) and 3 (Pearson's correlation = 0.77, $p = 1.34 \times 10^{-13}$). The correlation at passage 1 is somewhat lower (Pearson's correlation coefficient = 0.58, $p = 8.24 \times 10^{-7}$). This suggests that the

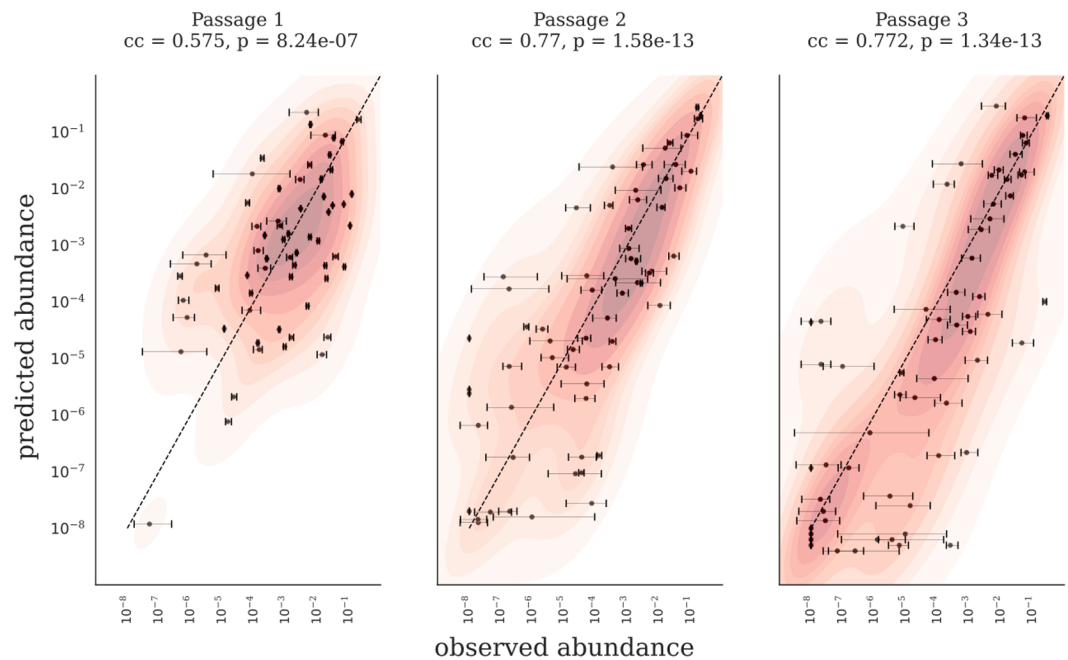


Fig 2. The model predicts serial dilution dynamics for a complex synthetic human gut microbiome. The model for hCom2 was validated by predicting strain abundances at different passages from inoculum abundances. We trained the model (Eq (3)) to fit R_i using strain abundances from passages 1 and 2 of the serial dilution experiment in Ref [31]. For validation, we used the fitted R_i to predict strain abundances in passages 1, 2, and 3, starting from inoculum abundances reported in the same study (Ref [31]). Pearson's correlation coefficients (cc) and p-values between \log_{10} of predicted and observed abundances are listed above each panel. Each point on the scatterplot represents one strain. Error bars correspond to the range (maximum minus minimum) of observed strain abundances across three biological replicates.

<https://doi.org/10.1371/journal.pcbi.1013222.g002>

model is better at predicting abundances at later passages, including steady state, than at the first passage.

Next, we evaluated our model's performance by comparing it against two null models: 1) monoculture null model and passage 2 null model. For the first null model, we assumed a non-interacting system where each strain was predicted to grow to its monoculture abundance using the inferred R_i . As expected, our model significantly outperformed this monoculture null model (compare Figs 2 and S12). The Pearson's correlation coefficients between predicted and observed abundances at steady state were 0.77 (p-value=2.5e-3) for our model and 0.35 (p-value=4.7e-3) for the monoculture null model. Furthermore, the RMSE on \log_{10} scale for the two models were 1.7 and 3.4, respectively.

For the second null model, we assumed that the community abundances at steady state (passage 3) were equal to the abundances at passage 2. This model does not capture the dynamics of the system, as it assumes a growth ratio (ratio of strain abundances at passages 3 and 2) equal to 1 for all strains, and consequently the observed and predicted growth ratios have zero covariance. In contrast, our model's predicted growth ratios showed a statistically significant correlation (Pearson's correlation coefficient = 0.46 on \log_{10} scale, $p = 1.33 \times 10^{-4}$) with the observed growth ratios. Additionally, the RMSE on \log_{10} scale between the predicted and observed growth ratios was 1.71 for our model, while it was 2.8 for the second null model.

To validate our model, we also conducted shuffling experiments on the consumption fluxes matrix ($n_S \times n_R$). We performed three types of shuffling: (1) complete randomization, (2) row sum-preserving shuffling, and (3) column sum-preserving shuffling (S14 Fig). Models trained on these shuffled consumption fluxes exhibited statistically significantly worse performance compared to the model trained on unshuffled fluxes (S14 Fig). This outcome suggests that our model's performance relies on biological information encoded in the consumption fluxes rather than the model's statistical flexibility. Furthermore, we observed a performance ranking among the shuffling experiments, in descending order: (1) column sum-preserved, (2) row sum-preserved, and (3) completely shuffled (S14 Fig). This ranking implies that preserving total consumption fluxes for metabolite clusters (columns) is more critical for achieving a good fit than preserving total consumption fluxes for individual strains (rows).

Accounting for variability in biological replicates. In addition to the correlation coefficient, Root Mean Squared Error (RMSE), on the \log_{10} scale, is another metric that can be used to quantify model performance. RMSE measures how much, on average, the predicted strain abundances deviate from the observed strain abundances on the log scale, and unlike correlation, a lower value for RMSE is desirable. S3 Fig shows that the RMSE between our model's predictions and observed abundances is between 2 (passages 2 and 3) and 3.6 (passage 1) times larger than replicate-to-replicate variability (RMSE between biological replicates). While RMSE of our prediction is sizeable, it is comparable to replicate-to-replicate variability.

Large biological replicate-to-replicate variability could result from undefined media [34], such as the mega media used in Ref [31] with significant variation in resource concentrations in the media between replicates and/or passages. Furthermore, in Ref [33], the same mega-media was used to measure consumption and production fluxes, resulting in significant day-to-day variation in $c_{\alpha i}$ and $p_{\gamma i}$ measurements. In addition to batch variability, other potential sources of replicate-to-replicate variation include differences in the lag phase following diauxic shift [16,35] and cell clumping [35]. Furthermore, technical errors introduced during DNA extraction or while executing DNA sequencing protocols can amplify this replicate-to-replicate variability [34]. These technical and biological sources can also contribute to the residual error in the model predictions.

The width of the error bars in Fig 2 appears to be larger for the low abundance strains compared to the high and medium abundance strains. We tested this quantitatively and found that the width of the error bars is indeed negatively correlated with strain abundances (S4d-S4f Fig). Another measure of the same effect is the RMSE between biological replicates, which also decreased steadily with increasing abundance (red curves in S4a-S4c Fig). Here, biological replicate-to-replicate variability was calculated as the cumulative RMSE between biological replicates over a subset of strains. We subset the strains by thresholding the average observed abundances at steady state (details are given in the Methods section).

In-silico experiments on the synthetic human gut microbial community

After validating our model, we used it to perform three in silico experiments to study the response of the synthetic gut community to different perturbations: In (i), we removed one strain from the inoculum and predicted the steady-state abundances of the remaining strains. By repeating this leave-one-out experiment for each of the 63 strains, we calculated the network of significant direct and indirect interactions between strains; In (ii), we increased the concentrations of individual resources 100-fold and examined the resulting changes in steady-state strain abundances. This experiment produced a matrix of direct and indirect interactions between resources and strains. Encouraged by experiments (i) and (ii), we tested our ability to manipulate the community through multi-nutrient supplementation. In (iii), we increased the concentrations of several resources to balance the abundances of strains as much as possible. The details of these three in silico experiments are described in the following sections.

Sensitivity of steady-state strain abundances to removal of a single strain from the inoculum. The first perturbation experiment we performed was to remove one strain from the inoculum and predict the steady-state abundances of the remaining strains. Each of 63 strains was removed in a different leave-one-out experiment. We compared the predicted steady-state abundances between the perturbed and unperturbed communities. This allowed us to estimate interactions between strains in the community, and the robustness of the community to extinction-driven perturbations. We found two types of interactions between strains - 1) cooperative (positive), and 2) competitive (negative). We represent these interactions as a directed graph in Fig 3a. For each edge, the source and destination nodes represent the removed and perturbed strains, respectively. To filter out noise, we kept only those edges where the abundance of the perturbed strain either increased more than 10-fold (red arrows in Fig 3a) or decreased more than 10-fold (green arrows in Fig 3a). Each edge is weighted by the absolute amount of change in abundance (on the \log_{10} scale) of the perturbed strain when the strain at the source of the edge is removed. Removing 11 of the 63 strains in independent leave-one-out experiments resulted in a significant shift in the abundance of at least one perturbed strain. Of the 63 strains, 35 were perturbed by the removal of at least one strain. Node sizes were set proportional to the weighted in-degree (sum of weights of all incoming edges at a node).

For cooperative interactions (green edges in Fig 3a), removal of a strain leads to a decrease in the abundance of the perturbed strain. One way this can happen is if the strain with decreased abundance cross-feeds on the metabolites produced by the removed strain. Consequently, when the strain is removed, the perturbed strain has fewer resources and responds with a decrease in abundance. To test this hypothesis, we calculated a cross-feeding score by counting the fraction of metabolites consumed by the perturbed strain that were produced by the removed strain (see Methods section for details on calculating the cross-feeding score). The score is equal to 1 when all metabolites produced by the removed strain are consumed by the perturbed strain and 0 when none of them are consumed. For cooperative or green edges,

we found that, the average cross-feeding score of 0.64 was statistically significantly higher ($p = 9.7 \times 10^{-5}$, two-tailed Mann-Whitney Wilcoxon test) than the average cross-feeding score of 0.55 between pairs of non-interacting strains or for strains connected by competitive interactions (red edges) in S7 Fig.

Next, we analyzed the competitive edges in the interaction network. In a competitive interaction, the removal of one strain leads to an increase in the abundance of another strain. In our model, the primary source of these interactions is competition for resources between the removed and perturbed strains. We calculated a competition score by counting the fraction of resources (or their clusters) consumed by the perturbed strain that were also consumed by the removed strain (see Methods section for details on calculating the competition score). For competitive edges, the average competition score of 0.65 between the removed and perturbed strains was significantly larger ($p = 0.02$, two-tailed Mann-Whitney Wilcoxon test) than the average competition score of 0.58 between non-interacting pairs or cooperative interactions (green edges) in S7 Fig.

The strain-strain interaction network's properties were analyzed using weighted out- and in-degree metrics from network analysis. We observed that strains with high out-degree, indicating a substantial impact on the community when removed, often correlated with high abundance in the unperturbed community (S5 Fig). This pattern persisted across both cooperative and competitive interactions (S5a and S5b Fig). The removal of a dominant strain typically results in resource reallocation, allowing competitors to expand and alter the community's relative abundance. Notably, this community shift is intrinsic and not an artifact of the normalization process in relative abundance studies, as demonstrated in S6a Fig, where the exclusion of the most abundant strain affects other strains by different amounts.

There were some exceptions to this global correlation between weighted out-degree and strain abundances. For instance, some intermediate abundance strains (*C. aerofaciens*, *B. eggerthii*, and *P. merdae*) disproportionately affected the community compared to their abundances (see S6 Fig). These strains could potentially be keystone species for the community.

We then examined the weighted in-degree distribution in the strain-strain interaction network, considering both cooperative and competitive edges. The total weighted in-degree of a strain represents its overall sensitivity to the removal of other strains. In Fig 3a, node sizes are scaled to their total weighted in-degree. Fig 3b and 3c display the average weighted in-degree, segregated by edge type (competitive vs cooperative), against the unperturbed abundance of strains at steady state. For competitive edges, there's a negative correlation between average in-degree and the \log_{10} of strain abundance (Pearson's cc of -0.27 , $p = 0.03$, Fig 3b). This trend is consistent with the resource conservation law, where the removal of a low abundance strain cannot possibly significantly affect higher abundance strains due to the limited resource reallocation it causes. The higher the abundance of the strain, the smaller the number of other strains that can potentially affect it, and thus the lower its in-degree should be. For several of the most abundant strains, we expect their in-degrees to be close to zero, as we see in Fig 3b.

Surprisingly, we found no significant correlation between weighted cooperative in-degree and strain abundance (Pearson's cc of 0.21 , $p = 0.1$, Fig 3c). Instead, according to our analysis, strains of intermediate abundance showed a higher tendency for cooperative interactions (Fig 3c). Notably, 5 out of 7 of these strains were from the genus *Bacteroides* (labeled in Fig 3a and 3c), representing a significant 71.4% occurrence compared to their 33.3% fraction in the total set of 63 strains ($p = 0.036$, one-sided hypergeometric test). This trend persists even at a lower cutoff, with *Bacteroides* comprising 9 of the top 15 most responsive strains ($p = 0.0154$, one-sided hypergeometric test). However, this pattern is limited to *Bacteroides* strains with intermediate abundances. In fact, the top 10 most abundant strains, including three

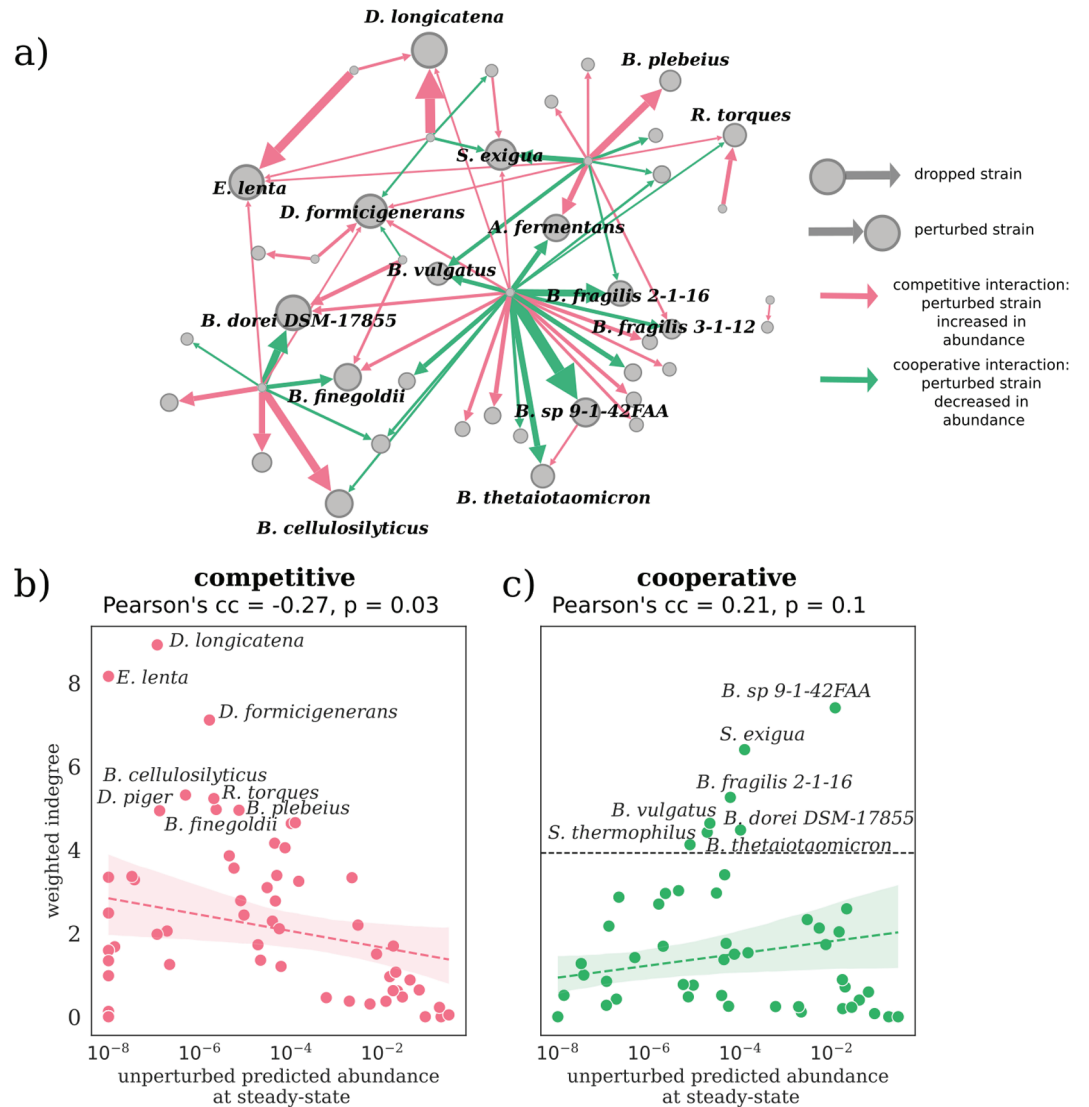


Fig 3. Leave-one-out perturbations reveal competitive and cooperative interactions between strains in the synthetic gut community. a) The directed strain-strain interaction network from leave-one-out experiments displays interaction edges directed from the removed strain to the perturbed strain and weighted by the log₁₀-scaled abundance change of perturbed strains. Node sizes reflect their total weighted in-degrees, with the top 15 strains labeled. Cooperative and competitive interactions are depicted as green and red edges, respectively. b) Weighted in-degrees of strains from competitive interactions (red) plotted against their steady-state abundances in the unperturbed community. The top 8 strains most sensitive to competitive interactions are labelled. c) Similar to b), but with in-degrees for cooperative interactions (green), with labels for the top 7 strains most sensitive to cooperative interactions. Panels b) and c) include Pearson's correlation coefficients between frequency and weighted degree, along with a dashed line fit and a shaded confidence interval.

<https://doi.org/10.1371/journal.pcbi.1013222.g003>

Bacteroides species (*Bacteroides uniformis* ATCC-15579, *Bacteroides thetaiotaomicron* VPI-5482, and *Bacteroides stercoris* ATCC-43183) have exactly zero weighted in-degree.

Sensitivity of the steady-state strain abundance to increase in the concentration of a single metabolite. Leave-one-out experiments show competition and cooperation between strains without identifying the responsible metabolites. To further understand this, we analyzed how the steady-state abundance of each strain changes when the concentration of a

single metabolite is increased. Since decreasing metabolite concentrations in undefined mega media is impossible in practice, we focused on sensitivity to increases. For each experiment, we increased the concentration of one of 38 metabolite clusters by a factor of 100. Then, using Eq (3), we predicted the new steady-state strain abundances. The logarithmic scale (base 10) difference between the perturbed and unperturbed abundances for each strain indicated its sensitivity to that metabolite.

Fig 4 presents a clustered heatmap illustrating how strain abundances respond to metabolite concentration perturbations. These perturbations caused both increases (33%) and decreases (67%) in strain abundances, revealing distinct clusters. For instance, strains *Coprococcus comes* ATCC 27758, *Holdemanella bififormis* DSM 3989, *Dorea longicatena* DSM 13814, and *Anaerotruncus colihominis* DSM 17241 showed notable abundance increases in response to several metabolites, possibly due to direct consumption or indirect effects like cross-feeding. Conversely, a cluster of mainly *Bacteroides* strains at the heatmap's top exhibited abundance decreases in response to 10 specific metabolites, likely from competition-related indirect effects.

Next, we also quantified the agreement between the strain-metabolite network inferred by our in silico experiment (Fig 4) and the consumption fluxes. For any given strain, we computed correlation between the vector of predicted log-fold change in the strain's abundance (in response to resource perturbations) and the vector of consumption fluxes for the same strain. We found that though the average correlation across strains was statistically significant, it was low with average Pearson's and Spearman's correlation coefficients being 0.23 (p-value = 7.57×10^{-20}) and 0.27 (p-value = 2.62×10^{-27}), respectively (S8 Fig). This low correlation can be explained by the fact that the strain-resource matrix from the resource perturbation experiments captures both direct and indirect interactions as shown in Fig 4. While the consumption and production fluxes cannot explain the indirect interactions without the strain-strain interaction network. For instance, many strains exhibit reduction in abundance in response to resource perturbations in Fig 4, which can only happen due to indirect interactions such as competition between strains for resources.

Some strains in Fig 4 remained largely unaffected by metabolite perturbations. For example, the following 10 strains marked with a black arrow in Fig 4 were perturbed by fewer than 3 metabolites: *Bacteroides caccae* DSM 43185, *Marvinbryantia formatexigens* DSM 14469, *Bacteroides coprophilus* DSM 18228, *Clostridium leptum* DSM 753, *Ruminococcus bromii* ATCC 8503, *Parabacteroides johnsonii* DSM 18315, *Lactococcus lactis* DSMZ 20729, *Lactobacillus ruminis* ATCC 25644, *Tyzzereella nexilis* DSM 2243, and *Catenibacterium mitsuokai* DSM 15897. Consequently, we excluded these strains from our subsequent in silico experiment, which focuses on equalizing steady-state strain abundances through increasing concentrations of multiple metabolites (detailed in the following section).

Supplementing the mega media with multiple metabolites can approximately equalize strain abundances. In the previous in silico experiment we saw that perturbations in individual metabolites can both increase and decrease strain abundances in hCom2*. This suggests that simultaneous perturbations in multiple metabolites may be able to, at least approximately, equalize strain abundances. Equalizing abundances has practical implications for gut microbiome transplantation therapy. If abundances can be equalized, then all strains in the transplanted community would have an equal footing to survive and colonize the gut.

To accomplish this we designed a greedy algorithm in which we changed abundances of multiple metabolites one-by-one. At every step we changed the abundance of a single metabolite that brings the community closest to uniformity. For this experiment we removed the 10 non-responsive strains identified in the previous section and marked with a black arrow

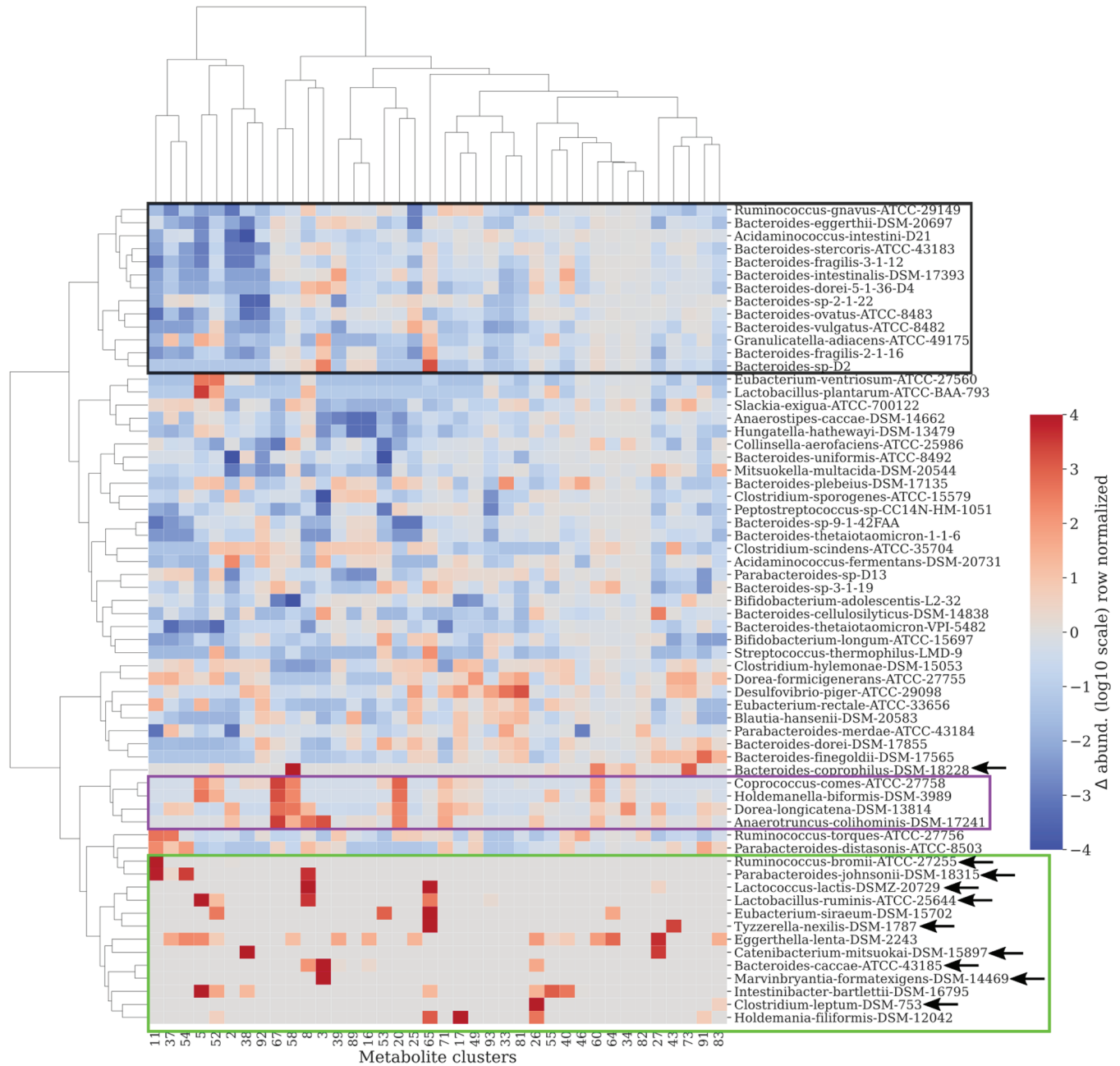


Fig 4. Response of strain abundances to increases in concentrations of individual metabolites. A hierarchically clustered heatmap of the ratio (on log₁₀ scale) between perturbed and unperturbed strain abundances at steady state in response to a 100-fold increase in the concentration of a single metabolite in the mega media. Columns and rows represent 38 metabolites (or their clusters) and 63 strains in the synthetic gut community, respectively. Three representative clusters of strains have been enclosed in boxes for easy identification. 10 strains marked with a black arrow were classified as poorly responsive to metabolite addition and were therefore excluded from the in silico experiment to equalize the strain abundances in Fig 5.

<https://doi.org/10.1371/journal.pcbi.1013222.g004>

in Fig 4. Details of the greedy algorithm are given in the Methods section. Since the greedy algorithm was stochastic, we repeated it 10 times to obtain 10 possible metabolite perturbation profiles and 10 different perturbed R_i .

We identified 56 metabolites that were altered in at least one of the 10 runs of the greedy algorithm. The abundances for these 56 metabolites are shown in a clustered heatmap in S9

Fig. 20 of the 38 non-zero unperturbed R_i are present in this list. The remaining metabolites had $R_i = 0$ before perturbation. We averaged the perturbed R_i over the 10 runs of the greedy algorithm and used this average to calculate the steady-state abundances starting from the inoculum for hCom2*. The community is now much closer to uniformity compared to the unperturbed scenario (Fig 5). The root mean square deviation (RMSD) (on \log_{10} scale) between the perturbed steady state and perfect equalization was 1.7, which was two times smaller than the \log_{10} -RMSD of 3.4 between the unperturbed steady state and perfect equalisation.

Discussion

Here, we introduced and studied a simplified consumer resource model for complex microbial communities with hundreds of coexisting strains growing on several hundred resources in serial dilution lab experiments. The central assumption of the model is that during the growth phase of the cycle, strains share resources in proportion to their average abundances multiplied by strain- and resource-specific consumption fluxes. This assumption was applied to all resources in the media to link strain abundances at successive passages of serial dilution experiments via mass conservation. We also incorporated cross-feeding into the model via a simple linear term linking the abundances of strains producing a given resource to its excess concentration in the medium. Consumption and production fluxes can be inferred from metabolomics experiments performed on batch growth experiments with individual strains. Our model can then be fitted to the data in a serial dilution experiment to infer resource concentrations, $R_i \geq 0$, by solving Eq (3) for a single dilution cycle as a non-negative least squares

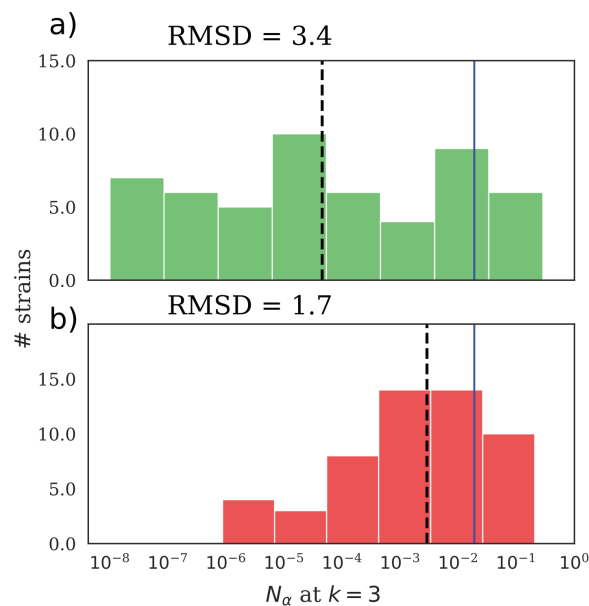


Fig 5. Equalisation of steady-state strain abundances for hCom2. Distribution of steady-state strain abundances for hCom2 before (a) and after (b) perturbation aimed at equalizing strain abundances. As a result of this perturbation, the \log_{10} -RMSD of the steady state abundances decreased twofold from 3.4 to 1.7. The blue solid lines show the perfectly equalized steady state abundance, the dashed black lines show the mean abundance values for the unperturbed and equalized communities.

<https://doi.org/10.1371/journal.pcbi.1013222.g005>

(NNLS) problem. With the fitted R_i , the model can be used to predict the dynamics at other dilution cycles not used in the initial fit.

We tested the model on a defined synthetic human gut microbiome, hCom2* growing on a rich medium with several hundred metabolites reaching steady-state diversity of around 60 strains [31]. To fit the model to hCom2*, we first obtained strain-specific resource consumption and production fluxes in the mega medium from the metabolomics experiments described in Han et al. 2021 [33]. The abundances of strains in hCom2* grown in mega medium at multiple passages of serial dilution experiments were obtained from Ref [31].

Modeling a complex synthetic community such as hCom2* requires several simplifying assumptions. First, our model assumes that multiple strains consuming the same resource share it in proportion to their average abundances during the growth cycle. An accurate calculation of the average abundance requires integration of the instantaneous abundance over the entire growth cycle, but these data were not available in Ref [31]. Since strains grow approximately exponentially, we approximated the average abundance by the geometric mean of the strain abundances at the beginning and end of the growth cycle, weighted by the time fraction f (see Eq (2)). In addition, depletion times will in principle differ between resources. This could be captured by making the time fraction resource specific. In principle, we could have fitted resource-specific time fractions f_i directly from the community dynamics. However, this was not practical with the limited training data we had. Therefore, we simplified our model to use the same time fraction for all resources. This simplification can be partially justified because hCom2 was constructed by augmenting a simpler community (hCom1) with species from a large pool. This process filled all ecological niches that remained open in hCom1 and placed the most competitive strains in each niche. We have previously shown that in such mature communities, most of the time during each growth cycle is spent in the first temporal niche where all resources are present [16]. After this long first niche, all resources rapidly disappear one after another. Therefore, both the differences in the depletion times of the resources and the deviations of the approximate average abundances from the exact time averages are expected to be relatively small. This a posteriori justifies our simplifying assumptions.

Another approximation of our model was the use of coarse-grained resource clusters, which was appropriate to describe the growth of ~ 60 species on nearly 300 resources. Our resource clustering strategy can be compared to previous work on consumer resource models (CRMs) for microbial communities of varying complexity. For low complexity communities, with just a few strains and resources, we and others have previously developed detailed CRMs [16,17,29,30], where each resource has its own separate depletion time. These differences in resource depletion times could in principle be captured in our model with a resource-specific time fraction f_i , but as explained above this was not feasible for the limited data we had. At the intermediate complexity level, Ho et al. 2022 [28] developed a CRM for a simplified synthetic human gut microbiome with 15 strains. They used binary consumption fluxes of these strains to group resources into $2^{15} = 32,768$ binary groups based on the exact subset of species capable of consuming them. They then retained about 30 of the most abundant groups. This method is not scalable to a more complex community like the hCom2* with 63 strains surviving in steady state. In fact, with binarized fluxes, there would be $2^{63} \sim 10^{19}$ possible binary metabolite groups, and searching this prohibitively large space is computationally infeasible. Therefore, we resorted to a more traditional approach of clustering metabolites with similar consumption and production fluxes, but not necessarily identical binary consumption profiles. Our model can be easily adapted to work with individual metabolites if a sufficient number of experiments is available to estimate each R_i individually. One way to accomplish this is to run additional serial dilution experiments with

inocula composed of subsets of strains of different diversity, in addition to the full diversity community where all strains are initially present.

Despite all the simplifications, our model performed reasonably well in predicting both the dynamics and the steady-state abundances of the strains in the serial dilution experiments of Ref [31]. More than 50% of the residual error in the model predictions was due to biological replicate-to-replicate variability in the serial dilution abundance data (S3 Fig). Experimentally, this replicate-to-replicate variability is most likely a consequence of variation in the composition of the mega medium - a rich, undefined medium. Variability in the composition of the mega-medium was also responsible for the day-to-day variation in consumption and production flux measurements observed in the experiments of Han et al. 2021 [33].

Using the model trained on the synthetic human gut microbiome hCom2*, we performed three in-silico perturbation experiments to study the organizational properties of the community. From the leave-one-out experiment (Fig 3b), we found that the intermediate abundance strains were the most sensitive to cooperative interactions (Fig 3c). Strains of the genus *Bacteroides* were clearly overrepresented in this group (5 out of 7 labeled strains in Fig 3c). This can be tentatively attributed to the fact that *Bacteroides* strains tend to be generalists [36,37], making them likely recipients of cooperative cross-feeding interactions, which in turn places them in the intermediate abundance tier of a multi-level trophic community. It should be noted that the nutrient composition of the mega-medium used in the in vitro experiments of Ref [31] is dramatically different from the polysaccharide-dominated environment of the human large intestine. Therefore, the trophic levels of *Bacteroides* strains in vivo [38,39] are likely to be different from those observed in vitro.

The second in silico experiment we performed was to perturb metabolite clusters individually by increasing their concentration, R_i , by a factor of 100 (Fig 4). The effect of resources on strains can be classified as either direct or indirect. For direct effects, strain abundance increases in response to an increase in the concentration of a metabolite that it consumes. Strain abundances may also increase due to indirect interactions such as cross-feeding. Decreases in strain abundance in response to increases in the concentration of a single metabolite can only occur through indirect interactions such as resource competition. Our model captures both direct and indirect effects of perturbations in metabolite concentrations. This is reflected in increased (33%) or decreased (67%) strain abundances in response to 100-fold increases in concentrations of individual metabolites.

In the third in silico experiment, we designed and implemented a greedy algorithm to equalize steady-state strain abundances. This approach may be relevant for practical applications such as gut microbiome transplantation or supplementation therapy. Indeed, the initial population densities of the members within a microbial community have been shown to influence the outcome of community assembly. In the absence of specific information, starting with equal initial densities for each strain is a logical approach, as it ensures all strains have an equal footing to colonize the recipient gut.

To enhance the practical applicability of our computational framework, the model can be extended to design and interpret mucin bead-based experiments described in Ref [31], where microbial colonization of mucosal surfaces is explicitly examined. By integrating strain-specific dilution factors that reflect differences in microbial adsorption onto mucin-coated beads, our model can predict which species are likely to successfully colonize and persist on mucosal surfaces under conditions that more closely mimic the human gut environment. Such predictions can practically inform the experimental design of gut microbiome transplants by identifying the most promising candidate strains.

In conclusion, we have introduced and studied a simplified consumer resource model capable of predicting the dynamics of a complex community of strains in a serial dilution

experiment. This model was tested on a defined synthetic human gut community consisting of a controlled collection of strains with known resource consumption and production fluxes. One of the potential future directions is to extend our model to other synthetic communities for which there is no metabolomics data to quantify consumption and production fluxes. One example of this with important practical applications is given by microbial strains isolated from the plant rhizosphere [40]. These strains were used to construct complex synthetic communities composed of 185 strains [14] and 62 strains [41] studied in *Arabidopsis thaliana* or 36 strains [13] studied in Sorghum. The application of our model to these communities would require a reliable way to predict of consumption and production fluxes of individual strains directly from their genomes. While computational methods cannot fully replace dedicated metabolomics experiments, they can be used as a first-order approximation. Promising approaches include in silico reconstruction of mechanistic genome-scale metabolic models (see [42] for a recent review) or “black box” machine learning algorithms to predict consumption and production of individual metabolites (see e.g. [43,44]).

Methods

Data

We parameterized our model for a synthetic human gut microbiome hCom2 [11] using two published and publicly available datasets. The first is a metabolomics dataset comprising consumption and production fluxes of 178 strains, including all hCom2 strains, for 833 metabolites, generated using an integrated liquid chromatography-mass spectrometry (LC-MS) pipeline described in Han et al. 2021 [33]. These strains were individually cultured in Mega Medium [45]-a rich, undefined medium known to support the growth of diverse bacteria. The culture supernatant was collected between mid-log and stationary phase for processing through the LC-MS pipeline. For each metabolite and strain, the consumption flux $c_{\alpha i}$ was calculated by subtracting from 1 the ratio of the concentrations of the metabolite i before and after batch growth of a single strain γ in the mega medium. Similarly, the total production flux $p'_{\gamma i}$ was calculated by subtracting 1 from the ratio of the metabolite i before and after batch growth of a single strain γ in the mega medium. The production flux $p_{\gamma i}$ used in Eq (3b) is calculated per unit biomass. It is given by the total production flux $p'_{\gamma i}$ divided by the total biomass N_{γ} (grown alone) of the strain γ at the end of the batch experiment. This biomass is computed as described below.

The second data set captures the dynamics of a synthetic human gut microbiome grown over 6 passages in a serial dilution experiment [31]. The set of 117 strains used in this experiment was extended from hCom2 [11]. In one type of experiment, these strains were grown in a medium containing different types of beads that provided surfaces for bacterial attachment. This experiment was designed to mimic the spatial organization of the human gut microbiome. The second type of experiment was a control using only the liquid Mega Medium without beads. Our model assumes an equal dilution ratio of each strain and is only suitable for the control experiment without beads. In principle, it is possible to adapt our model to describe other experiments in Ref [31], where passage from one growth cycle to the next is achieved by transferring a single bead. However, it requires 63 new parameters that quantify the degree of adhesion of each of the strains to the beads. It is not computationally feasible to fit these parameters without additional experiments, so we limited our study to no-bead control experiments. Cultures were grown in Mega Medium [45] for three days and then passaged with the dilution factor $D = 15,000$. Already after three serial passages the community was observed to reach the steady state [31]. Therefore, we limited our model to describe the community dynamics during the first three passages. Each serial passaging experiment was

run in three biological replicates, with each biological replicate after each passage sequenced in three technical replicates. For our analysis, we averaged the abundances from the technical replicates. Our model was necessarily limited to include only the 63 strains from Ref [31] for which we had consumption and production fluxes from Ref [33]. However, these strains accounted for $\sim 90\%$ of the total abundance of all strains surviving in the steady state and thus provided a reasonably good approximation to the full community studied in Ref [31]. The names of these strains, along with the abbreviations we assigned to them, are listed in S1 Table.

Clustering metabolites

The number of strains (63) used in our model is much smaller than the initial number of individual metabolites (292) in the Mega Medium. We clustered the metabolites using the consumption fluxes. The clustering procedure gave us 10 non-singleton and 88 singleton metabolite clusters. These 98 clusters were used for all analyses in the paper. The names of the metabolites in each metabolite cluster are given in S2 Table. In the metabolomics data from Han et al. 2021 [33], some metabolite names appear multiple times. This occurs because Ref [33] reported spectra from the same compound collected using different analytical methods separately. We observed significant differences in consumption and production fluxes for these compounds across methods. Therefore, we preserved these repeats by adding numerical suffixes.

For clustering we first binarized the consumption fluxes $c_{\alpha i}$ with a threshold of 0.3, and used only those metabolites consumed by more than 5 strains. The 0.3 threshold maximizes Spearman's rank correlation between steady-state strain abundances and their degree (number of consumed resources) in the binarized consumption fluxes (see S13 Fig). The metabolites selected after binarization were grouped into 10 non-singleton clusters using hierarchical agglomerative clustering with Euclidean distance as the metric and Ward's linkage method. The remaining metabolites consumed by less than 5 strains, but at least one strain, were used as singleton clusters. In total, we obtained 98 metabolite clusters comprising a total of 224 metabolites.

To get $c_{\alpha i}$ for a non-singleton cluster i , we first took the average over all metabolites j in this cluster: $\bar{c}_{\alpha i} = -\sum_j \log(1 - c_{\alpha j}) / (\# \text{ of metabolites in the cluster})$. Then, $c_{\alpha i}$ for the cluster was obtained by the inverse transformation $1 - e^{-\bar{c}_{\alpha i}}$.

Similarly, for the total metabolite $p'_{\gamma i}$ for the non-singleton cluster i , we took the average $\bar{p}'_{\gamma i} = \sum_j \log(1 + p'_{\gamma j}) / (\# \text{ of metabolites in the cluster})$. Then, the production flux for the cluster was obtained by the inverse transformation $p'_{\gamma i} = e^{\bar{p}'_{\gamma i}} - 1$.

Estimation of R_i for metabolite clusters

R_i were estimated from Eq (3) using $c_{\alpha i}$, $p_{\gamma i}$, and N_{α} (end of cycle) as the inputs and with $f = 0.9$, which was estimated using our strategy detailed in one of the following sections. The strain abundances were used only for the passages 1 and 2. These two passages were the most dynamic for hCom2 [31]. Passages 1 and 2 were used as N_{α} (end of cycle 1) and N_{α} (end of cycle 2), respectively. Eq (3) was applied to each biological replicate. This resulted in a system of linear equations in R_i , which was solved using non-negative least squares (NNLS) [46] to obtain an estimate of non-negative R_i . This NNLS problem was solved using the `nnls` function from the `optimize` subpackage of the SciPy Python package [47].

The production flux $p_{\gamma i}$ used in Eq (3b) is given by the total metabolite production flux $p'_{\gamma i}$ derived from the metabolomics experiment [33] normalized by the biomass of this strain

N_γ (grown alone) in the end of the batch experiment: $p_{\gamma i} = p'_{\gamma i}/N_\gamma$ (grown alone). In practice, we do not know these abundances but can estimate them using the mass conservation: N_γ (grown alone) = $\sum_i R_i [c_{\gamma i} > 0]$ ($[\]$ is the Iverson bracket), which in turn depends on R_i . We used an iterative approach to jointly solve Eq (3) for R_i and N_γ (grown alone). First, we assigned N_γ (grown alone) = $1/n_s$. Next, we inferred R_i and used this value to update our estimate for N_γ (grown alone) using the aforementioned approximation. The updated N_γ (grown alone) was used to obtain the next estimate for R_i . This iterative scheme was repeated for 100 iterations. Empirically, we found that the iterative approach always converged in less than 100 iterations. The values of N_γ (grown alone) and R_i obtained at the end of the procedure were used for downstream analyses.

Predicting strain abundances from the model

With R_i obtained from the NNLS fit as described above, we used Eq (3) to predict the strain abundances forward in time. This was done starting with the experimentally measured strain abundances in the inoculum (averaged over two biological replicates) as N_α (end of cycle 0). Given R_i , $c_{\alpha i}$, $p_{\gamma i}$, N_α (end of cycle 0) and f (see the next section on how we fitted f), Eqs (3) for each α fully define N_α (end of cycle 1) (the number of equations is equal to the number of unknowns). We solved this system of nonlinear equations iteratively. First, N_α (end of cycle 1) were set equal to N_α (end of cycle 0) on the right-hand side (RHS) of Eq (3). The RHS was then used to calculate the new estimate for N_α (end of cycle 1) on the left-hand side (LHS). This new estimate was again substituted into the RHS to obtain an updated estimate for N_α (end of cycle 1). This was continued for 100 iterations to obtain the final estimate for N_α (end of cycle 1).

There were two biological replicates for the inoculum, which were not matched to the three biological replicates of serial passage experiments. Therefore, to predict strain abundances, we averaged the two inoculum abundances and used that as the starting point. As a result the model generates only one prediction for all passages, whereas the observed data had three different biological replicates. We took the geometric average (on \log_{10} scale) of the three biological replicates to compare against the model prediction. The variation in the three biological replicates is shown as horizontal error bars in Fig 2.

Estimation of the time fraction f

To fit the time fraction parameter f for hCom2, we estimated R_i for linearly spaced values of f in the interval $[0.3, 1]$ as shown in S1 Fig. For each estimated R_i , we made a prediction for the steady-state strain abundances reached at the third passage starting from the inoculum. The predicted abundances were compared to the observed abundances. As mentioned before, we took the geometric average (on \log_{10} scale) of the three biological replicates to compare against the model prediction. The value of f that gave the best agreement with the observed data was selected. The quality of the agreement was quantified by the Pearson's correlation coefficient between the logarithms of the predicted and observed abundances averaged over three biological replicates. Using this approach we found a sharp peak around $f = 0.9$ for hCom2 (S1 Fig). Therefore, $f = 0.9$ was used in our model throughout this study.

Estimate of replicate-to-replicate variability in experimentally observed abundances

For each passage, except the inoculum, three biological replicates were provided for the abundance of each strain [31]. To estimate the replicate-to-replicate variability at each

passage, for a given pair of replicates, Root Mean Square Error (RMSE) was calculated using \log_{10} of strain abundances. This was repeated for the 3 possible replicate pairs. The average RMSE for the 3 pairs of biological replicates was used as an estimate of the replicate-to-replicate variability.

Estimation of the cumulative RMSE over steady-state abundances

To calculate the cumulative \log_{10} RMSE for either the predictive performance of the model or the replicate-to-replicate variability in [S3a-S3c Fig](#), the strains were sorted in an increasing order of observed steady-state abundances (averaged over biological replicates). We then considered a set of different abundance thresholds (evenly spaced on the \log_{10} scale). For each threshold, all strains above that threshold were retained and the others were dropped. The RMSE of \log_{10} of the strain abundances was calculated for the subset of strains exceeding a given threshold.

Average competition and the cross-feeding scores between strains

For each pair of bacterial strains, we calculate a 'competition score'. This score shows how much they compete for the same food resources, which in this case are metabolites or metabolite clusters. We only look at metabolites that are present in the system (those with a non-zero concentration R_i). To make this score easy to understand, we rescale it to a scale from 0 to 1. This is done by dividing the number of shared metabolites by the total number of metabolites consumed by the perturbed strain. We average this score across all strain pairs to get an overall competition score.

Similarly, we calculate a 'cross-feeding score'. This score measures how many metabolites produced by one strain are consumed by another strain. This score is also normalized to be between 0 and 1 by dividing it by the number of metabolites consumed by the perturbed strain.

Greedy algorithm to equalize strain abundances

At each iteration of our algorithm, we consider the effect of increasing or decreasing the concentration of a single metabolite R_i on the strain abundances and choose a perturbation that brings them closest to uniformity. When increasing the concentration, we simply multiply R_i by 10. Metabolites initially absent in the medium are assigned a nominal very low concentration $R_i = 10^{-8}$. Since we do not allow the concentration of any metabolite to fall below its unperturbed value in the Mega Medium, we have implemented a special rule for decreasing the concentration, summarized in the following equation:

$$\log_{10} R_i(\text{Next Step}) = \log_{10} R_i(\text{Mega Medium}) + 0.6 \times (\log_{10} R_i(\text{Current Step}) - \log_{10} R_i(\text{Mega Medium})) \quad (4)$$

This rule ensures that the recipe discovered by our algorithm can be implemented experimentally by supplementing the Mega Medium with prescribed concentrations of selected metabolites. Indeed, there is no practical way to remove individual metabolites from a complex undefined medium such as the Mega Medium. At each iteration, for each R_i , a fair coin toss decides whether the concentration is increased or decreased, which adds stochasticity to the algorithm. The greedy step of our algorithm is repeated up to 500 times or until no further improvements can be made (local uniformity optimum).

The algorithm was repeated 10 times to obtain 10 possible perturbed R_i 's, which were then processed as described in the Results section.

Supporting information

S1 Fig. Estimating the time fraction f for hCom2. The time fraction f was estimated by inferring R_i from our model for linearly spaced values of f in the interval $[0.3, 1]$. For each estimated R_i , we made a prediction of the steady-state strain abundances starting from the inoculum using our model. Pearson's correlation coefficients between predicted and observed abundances for different values of f are plotted.

(PDF)

S2 Fig. Clustered heatmap of the consumption fluxes for the 98 metabolite clusters for hCom2. Columns and rows represent the metabolite clusters and the strains, respectively. From the left, the first 10 columns represent the non-singleton clusters, followed by the 88 singletons.

(PDF)

S3 Fig. Biological replicate-to-replicate variability explains more than half of the residual error in our model predictions for hCom2. Biological replicate-to-replicate variability at each passage was estimated as the root mean squared error (RMSE) between the strain abundances for the biological replicates. Model performance at each passage was estimated as the RMSE between predicted and observed strain abundances. The adjusted RMSE was then calculated as the difference between the RMSE for model performance and the RMSE for biological replicate-to-replicate variability.

(PDF)

S4 Fig. Biological replicate-to-replicate variability and model performance are negatively correlated with strain abundances in hCom2. a)–c) Cumulative \log_{10} RMSE is plotted as a function of average observed strain abundances at steady state for model performance (blue curve) and biological replicate-to-replicate variability (red curve). First, the strains were sorted in increasing order of average observed steady-state abundances. For each value of observed steady-state abundance used as a threshold, the cumulative RMSE at each passage was calculated for a subset of strains with strain abundances at that passage greater than the threshold (details are given in the main Methods section). d)–f) The width of the error bars (maximum minus minimum) from Fig. 2 are plotted as a function of observed strain abundance for different passages. Each point on the scatterplot represents a single strain. Pearson's correlation coefficient (cc) (along with the p -value) is given for the different passages. The linear regression fit to the scatterplot is also shown as a dashed line, with the 95% confidence interval shown as a shaded area around the linear fit.

(PDF)

S5 Fig. Weighted strain out-degree versus steady-state abundance for hCom2. Total weighted out-degree restricted to competitive (red, a) and cooperative (red, b) edges in the strain-strain interaction network for all the strains plotted against predicted strain abundances at steady state (passage 3) for the unperturbed hCom2 community. Some intermediate abundance strains have been highlighted in the plots. These strains have a disproportionately large impact on the community compared to their abundances in the unperturbed community.

(PDF)

S6 Fig. Perturbations from leave-one-experiment are non-trivial and cannot be explained by re-normalization for relative abundances for hCom2. a) Removal of the highest abundance strain from hCom2 causes steady-state abundances for multiple strains to change by different factors, which cannot be explained by re-normalization of perturbed abundances. b) Removal of a low abundance strain does not change steady-state abundances for the hCom2 community.

(PDF)

S7 Fig. Competition and cross-feeding underlie competitive and cooperative edges for hCom2. (Left) Comparison of competition scores for competitive edges (red) against random edges. Non-interacting and cooperative edges (green) are defined as random in this context. We obtained a mean competition score of 0.65 for competitive edges vs. 0.58 for non-interacting or cooperative interaction edges ($p = 0.02$, two-tailed Mann-Whitney Wilcoxon test). (Right) Comparison of cross-feeding scores for cooperative edges (green) against random edges. Non-interacting and competitive edges (red) are defined as random in this context. We obtained a mean cross-feeding score of 0.64 for cooperative edges vs. 0.55 for non-interacting or competitive interaction edges ($p = 9.7 \times 10^{-5}$, two-tailed Mann-Whitney Wilcoxon test).

(PDF)

S8 Fig. Histogram of Pearson's and Spearman's correlation coefficients across strains. Histogram across strains of a) Pearson's and b) Spearman's correlation coefficients between the predicted log-fold change in strain abundances (in response to resource perturbations) and the consumption fluxes. The black solid curves represent the kernel density estimates. The red dashed lines show the average values.

(PDF)

S9 Fig. Hierarchically clustered heatmap of R_i perturbed in-silico. Hierarchically clustered heatmap of the 10 perturbed R_i solutions generated by our greedy algorithm for equalising strain abundances at steady-state for hCom2. The columns and rows represent the different solutions and metabolite clusters, respectively.

(PDF)

S10 Fig. Clustered heatmap of the consumption fluxes for 224 metabolites for hCom2.

We have only included the metabolites which contributed to the biomass of at least one strain after thresholding the consumption matrix as described in the methods section. Columns and rows represent the metabolite numbers and the strains, respectively. Metabolites are clustered such that the 10 non-singleton metabolite clusters (their individual metabolites) first appear from the left, followed by the metabolites in the non-singleton clusters. The mapping between metabolite IDs on the x-axis and metabolite names are given in S3 Table.

(PDF)

S11 Fig. Heatmap of the production fluxes for 224 metabolites for hCom2. We have only included the metabolites which contributed to the biomass of at least one strain after thresholding the consumption matrix as described in the methods section. Columns and rows represent the metabolite numbers and the strains, respectively. Since the production fluxes spanned several orders of magnitude, they were log-transformed after addition of a 1 in the heatmap. Metabolites are clustered such that the 10 non-singleton metabolite clusters (their individual metabolites) first appear from the left, followed by the metabolites in the non-singleton clusters. The mapping between metabolite IDs on the x-axis and metabolite names are given in S3 Table.

(PDF)

S12 Fig. Predictions from the monoculture null model. Predictions from the monoculture null model for hCom2 plotted against observed abundances at different passages. Pearson's correlation coefficients (cc) and p-values between \log_{10} of predicted and observed abundances are listed above each panel. RMSE values are also shown. Each point on the scatterplot represents one strain. Error bars correspond to the range (maximum minus minimum) of observed strain abundances across three biological replicates.

(PDF)

S13 Fig. Selection of threshold for binarizing consumption fluxes before clustering metabolites into clusters. The plot illustrates Spearman's rank correlation between steady-state strain abundances and their degree (total consumed resources) in binarized consumption fluxes, as a function of consumption threshold. A global maximum occurs at a threshold of 0.3(30%).

(PDF)

S14 Fig. Effect of randomizing the consumption fluxes. The plot depicts a) Pearson's correlation coefficient and b) RMSE between predicted and observed steady-state (passage 3) strain abundances in the community on \log_{10} scale. We conducted experiments with three types of shuffling of the consumption fluxes matrix ($n_S \times n_R$): (1) complete randomization (shuffled), (2) row sum-preserving shuffling, and (3) column sum-preserving shuffling. The solid black line shows the performance for the model trained with the unshuffled consumption fluxes.

(PDF)

S1 Text. Approximate analytical estimate for time fraction f .

(PDF)

S1 Table. Abbreviations for strain names.

(PDF)

S2 Table. Metabolite clusters sorted in decreasing order of R_i for hCom2.

(PDF)

S3 Table. Mapping between metabolite IDs and metabolite names.

(CSV)

Acknowledgments

We thank Veronika Dubinkina and Akshit Goyal for useful discussions.

Author contributions

Conceptualization: Tarun Mahajan, Sergei Maslov.

Data curation: Tarun Mahajan, Sergei Maslov.

Formal analysis: Tarun Mahajan, Sergei Maslov.

Funding acquisition: Sergei Maslov.

Methodology: Tarun Mahajan, Sergei Maslov.

Project administration: Sergei Maslov.

Resources: Sergei Maslov.

Software: Tarun Mahajan.

Supervision: Sergei Maslov.

Validation: Tarun Mahajan.

Visualization: Tarun Mahajan.

Writing – original draft: Tarun Mahajan, Sergei Maslov.

Writing – review & editing: Tarun Mahajan, Sergei Maslov.

References

1. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012;489(7415):220–30. <https://doi.org/10.1038/nature11550> PMID: 22972295
2. Jovel J, Dieleman LA, Kao D, Mason AL, Wine E. The human gut microbiome in health and disease. *Metagenomics*. Elsevier; 2018. p. 197–213. <https://doi.org/10.1016/b978-0-08-102268-9.00010-0>
3. Marchesi JR. Prokaryotic and eukaryotic diversity of the human gut. *Adv Appl Microbiol*. 2010;72:43–62. [https://doi.org/10.1016/S0065-2164\(10\)72002-5](https://doi.org/10.1016/S0065-2164(10)72002-5) PMID: 20602987
4. Torsvik V, Øvreås L. Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol*. 2002;5(3):240–5. [https://doi.org/10.1016/s1369-5274\(02\)00324-7](https://doi.org/10.1016/s1369-5274(02)00324-7) PMID: 12057676
5. Garbeva P, Van Veen JA, Van Elsas JD. Microbial diversity in soil: selection of microbial populations by plant and soil type and implications for disease suppressiveness. *Annu Rev Phytopathol*. 2004;42:243–70. <https://doi.org/10.1146/annurev.phyto.42.012604.135455>
6. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci U S A*. 2006;103(32):12115–20. <https://doi.org/10.1073/pnas.0605127103> PMID: 16880384
7. Das S, Lyla P, Khan SA. Marine microbial diversity and ecology: importance and future perspectives. *Curr Sci*. 2006;:1325–35.
8. Santelli CM, Orcutt BN, Banning E, Bach W, Moyer CL, Sogin ML, et al. Abundance and diversity of microbial life in ocean crust. *Nature*. 2008;453(7195):653–6. <https://doi.org/10.1038/nature06899> PMID: 18509444
9. Nelson MC, Morrison M, Yu Z. A meta-analysis of the microbial diversity observed in anaerobic digesters. *Bioresour Technol*. 2011;102(4):3730–9. <https://doi.org/10.1016/j.biortech.2010.11.119> PMID: 21194932
10. Li L, He Q, Ma Y, Wang X, Peng X. Dynamics of microbial community in a mesophilic anaerobic digester treating food waste: relationship between community structure and process stability. *Bioresour Technol*. 2015;189:113–20. <https://doi.org/10.1016/j.biortech.2015.04.015> PMID: 25879178
11. Cheng AG, Ho PY, Jain S, Feiqiao BY, Meng X, et al. Design, construction, and in vivo augmentation of a complex gut microbiome. *Cell*. 2022;185(19):3617–36.
12. Vrancken G, Gregory AC, Huys GRB, Faust K, Raes J. Synthetic ecology of the human gut microbiota. *Nat Rev Microbiol*. 2019;17(12):754–63. <https://doi.org/10.1038/s41579-019-0264-8> PMID: 31578461
13. Chai YN, Ge Y, Stoerger V, Schachtman DP. High-resolution phenotyping of sorghum genotypic and phenotypic responses to low nitrogen and synthetic microbial communities. *Plant Cell Environ*. 2021;44(5):1611–26. <https://doi.org/10.1111/pce.14004> PMID: 33495990
14. Finkel OM, Salas-González I, Castrillo G, Spaepen S, Law TF, Teixeira PJPL, et al. The effects of soil phosphorus content on plant microbiota are driven by the plant phosphate starvation response. *PLoS Biol*. 2019;17(11):e3000534. <https://doi.org/10.1371/journal.pbio.3000534> PMID: 31721759
15. Fridman Y, Wang Z, Maslov S, Goyal A. Fine-scale diversity of microbial communities due to satellite niches in boom and bust environments. *PLoS Comput Biol*. 2022;18(12):e1010244. <https://doi.org/10.1371/journal.pcbi.1010244> PMID: 36574450
16. Wang Z, Goyal A, Dubinkina V, George AB, Wang T, Fridman Y, et al. Complementary resource preferences spontaneously emerge in diauxic microbial communities. *Nat Commun*. 2021;12(1):6661. <https://doi.org/10.1038/s41467-021-27023-y> PMID: 34795267
17. Bloxham B, Lee H, Gore J. Biodiversity is enhanced by sequential resource utilization and environmental fluctuations via emergent temporal niches. *PLoS Comput Biol*. 2024;20(5):e1012049. <https://doi.org/10.1371/journal.pcbi.1012049>

18. Erez A, Lopez JG, Weiner BG, Meir Y, Wingreen NS. Nutrient levels and trade-offs control diversity in a serial dilution ecosystem. *Elife*. 2020;9:e57790. <https://doi.org/10.7554/eLife.57790> PMID: 32915132
19. Lotka AJ. *Elements of physical biology*. Williams & Wilkins; 1925.
20. Volterra V. Variations and fluctuations of the number of individuals in animal species living together. *ICES J Marine Sci*. 1928;3(1):3–51. <https://doi.org/10.1093/icesjms/3.1.3>
21. Venturelli OS, Carr AC, Fisher G, Hsu RH, Lau R, Bowen BP, et al. Deciphering microbial interactions in synthetic human gut microbiome communities. *Mol Syst Biol*. 2018;14(6):e8157. <https://doi.org/10.15252/msb.20178157> PMID: 29930200
22. Momeni B, Xie L, Shou W. Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions. *Elife*. 2017;6:e25051. <https://doi.org/10.7554/eLife.25051> PMID: 28350295
23. MacArthur R, Levins R. Competition, habitat selection, and character displacement in a patchy environment. *Proc Natl Acad Sci U S A*. 1964;51(6):1207–10. <https://doi.org/10.1073/pnas.51.6.1207> PMID: 14215645
24. MacArthur R. Species packing and competitive equilibrium for many species. *Theor Popul Biol*. 1970;1(1):1–11. [https://doi.org/10.1016/0040-5809\(70\)90039-0](https://doi.org/10.1016/0040-5809(70)90039-0) PMID: 5527624
25. Chesson P. MacArthur's consumer-resource model. *Theor Popul Biol*. 1990;37(1):26–38. [https://doi.org/10.1016/0040-5809\(90\)90025-q](https://doi.org/10.1016/0040-5809(90)90025-q)
26. Tilman D. Resources: a graphical-mechanistic approach to competition and predation. *Am Naturalist*. 1980;116(3):362–93. <https://doi.org/10.1086/283633>
27. Tilman D. *Resource competition and community structure*. Princeton University Press; 1982.
28. Ho PY, Nguyen TH, Sanchez JM, DeFelice BC, Huang KC. Resource competition predicts assembly of gut bacterial communities in vitro. *Nat Microbiol*. 2024:1–13. <https://doi.org/10.1038/s41564-024-01625-w>
29. Wang Z, Fu Y, Goyal A, Maslov S. Fitness advantage of sequential metabolic strategies emerges from community interactions in strongly fluctuating environments bioRxiv, 2024:2024-06. <https://doi.org/10.1101/2024.06.14.599039>
30. Bloxham B, Lee H, Gore J. Diauxic lags explain unexpected coexistence in multi-resource environments. *Mol Syst Biol*. 2022;18(5):e10630. <https://doi.org/10.15252/msb.202110630> PMID: 35507445
31. Jin X, Yu FB, Yan J, Weakley AM, Dubinkina V, Meng X, et al. Culturing of a complex gut microbial community in mucin-hydrogel carriers reveals strain- and gene-associated spatial organization. *Nat Commun*. 2023;14(1):3510. <https://doi.org/10.1038/s41467-023-39121-0> PMID: 37316519
32. Connors BM, Thompson J, Ertmer S, Clark RL, Pflieger BF, Venturelli OS. Control points for design of taxonomic composition in synthetic human gut communities. *Cell Syst*. 2023;14(12):1044–1058.e13. <https://doi.org/10.1016/j.cels.2023.11.007> PMID: 38091992
33. Han S, Van Treuren W, Fischer CR, Merrill BD, DeFelice BC, Sanchez JM, et al. A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature*. 2021;595(7867):415–20. <https://doi.org/10.1038/s41586-021-03707-9> PMID: 34262212
34. Nearing JT, Comeau AM, Langille MGI. Identifying biases and their potential solutions in human microbiome studies. *Microbiome*. 2021;9(1):113. <https://doi.org/10.1186/s40168-021-01059-0> PMID: 34006335
35. Pacciani-Mori L, Giometto A, Suweis S, Maritan A. Dynamic metabolic adaptation can promote species coexistence in competitive microbial communities. *PLoS Comput Biol*. 2020;16(5):e1007896. <https://doi.org/10.1371/journal.pcbi.1007896> PMID: 32379752
36. Ryan D, Prezza G, Westermann AJ. An RNA-centric view on gut Bacteroidetes. *Biol Chem*. 2020;402(1):55–72. <https://doi.org/10.1515/hsz-2020-0230> PMID: 33544493
37. Louis P. Different substrate preferences help closely related bacteria to coexist in the gut. *mBio*. 2017;8(6):e01824–17. <https://doi.org/10.1128/mBio.01824-17> PMID: 29114031
38. Wang T, Goyal A, Dubinkina V, Maslov S. Evidence for a multi-level trophic organization of the human gut microbiome. *PLoS Comput Biol*. 2019;15(12):e1007524. <https://doi.org/10.1371/journal.pcbi.1007524> PMID: 31856158
39. Goyal A, Wang T, Dubinkina V, Maslov S. Ecology-guided prediction of cross-feeding interactions in the human gut microbiome. *Nat Commun*. 2021;12(1):1335. <https://doi.org/10.1038/s41467-021-21586-6> PMID: 33637740
40. Beck AE, Kleiner M, Garrell A-K. Elucidating plant-microbe-environment interactions through omics-enabled metabolic modelling using synthetic communities. *Front Plant Sci*. 2022;13:910377. <https://doi.org/10.3389/fpls.2022.910377> PMID: 35795346

41. Carlström CI, Field CM, Bortfeld-Miller M, Müller B, Sunagawa S, Vorholt JA. Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. *Nat Ecol Evol.* 2019;3(10):1445–54. <https://doi.org/10.1038/s41559-019-0994-z> PMID: 31558832
42. Fang X, Lloyd CJ, Palsson BO. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat Rev Microbiol.* 2020;18(12):731–43. <https://doi.org/10.1038/s41579-020-00440-4> PMID: 32958892
43. Gowda K, Ping D, Mani M, Kuehn S. Genomic structure predicts metabolite dynamics in microbial communities. *Cell.* 2022;185(3):530-546.e25. <https://doi.org/10.1016/j.cell.2021.12.036> PMID: 35085485
44. Gralka M, Pollak S, Cordero OX. Genome content predicts the carbon catabolic preferences of heterotrophic bacteria. *Nat Microbiol.* 2023;8(10):1799–808. <https://doi.org/10.1038/s41564-023-01458-z> PMID: 37653010
45. Romano KA, Vivas EI, Amador-Noguez D, Rey FE. Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide. *mBio.* 2015;6(2):e02481. <https://doi.org/10.1128/mBio.02481-14> PMID: 25784704
46. Lawson CL, Hanson RJ. Solving least squares problems. SIAM; 1995.
47. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* 2020;17(3):261–72. <https://doi.org/10.1038/s41592-019-0686-2> PMID: 32015543