Metabolic Model-based Ecological Modeling for Probiotic Design

James D. Brunner$^{1,2*}$ and Nicholas Chia$^3*$

$^1$Biosciences Division, Los Alamos National Laboratory, Los Alamos, NM, USA; $^2$Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA; $^3$Data Science and Learning, Argonne National Laboratory, Lemont, IL, USA

Abstract The microbial community composition in the human gut has a profound effect on human health. This observation has led to extensive use of microbiome therapies, including over-the-counter “probiotic” treatments intended to alter the composition of the microbiome. Despite so much promise and commercial interest, the factors that contribute to the success or failure of microbiome-targeted treatments remain unclear. We investigate the biotic interactions that lead to successful engraftment of a novel bacterial strain introduced to the microbiome as in probiotic treatments. We use pairwise genome-scale metabolic modeling with a generalized resource allocation constraint to build a network of interactions between taxa that appear in an experimental engraftment study. We create induced sub-graphs using the taxa present in individual samples and assess the likelihood of invader engraftment based on network structure. To do so, we use a generalized Lotka-Volterra model, which we show has strong ability to predict if a particular invader or probiotic will successfully engraft into an individual's microbiome. Furthermore, we show that the mechanistic nature of the model is useful for revealing which microbe-microbe interactions potentially drive engraftment.

Introduction Microbiome research has come to encompass key areas of disease, ranging from infections (Anantharam et al., 2013; Honda and Littman, 2012; Battaglioli et al., 2018) and cancer prevention (Moss and Blaser, 2005; Walther-António et al., 2016; Kim et al., 2020) to systemic immune and neurological responses (Severance et al., 2016; Kang et al., 2014; Chen et al., 2016). The effect of the microbiome on health is now undeniable, and every year in the US over 400,000 people collectively spend $1 billion dollars on over-the-counter probiotics intended to alter their microbiome (Kristensen et al., 2016). Many of the purported interactions between microbes and health involve resident microbiota and their interactions with the host, i.e., the interface between microbial ecology and human health. The goal of microbiome-targeted interventions is therefore to promote health by “restoring and maintaining the microbiota and the crucial health-associated ecosystem services that it provides” (Costello et al., 2012).

Despite the many links between the microbiome and health, our ability to deploy probiotics to modify the microbiome as intended has been met with relatively little success (Mullard, 2016; Zhu et al., 2019; Yuan et al., 2017; Zhao et al., 2021; Wu et al., 2017; Lawson et al., 2019). Studies looking at the ecological effects of probiotic administration show that administration of a probiotic is not sufficient to alter the community in the desired way. Specifically, engraftment of the administered microbial species is often limited, with only one-third to one-half of patients showing any signs of medium or long-term engraftment (Maldonado-Gómez et al., 2016; Pudgar et al., 2021). We offer the argument that probiotic interventions are primarily ecologic in nature; their purpose is to
While potentially predictive, such classifier approaches are sensitive to the underlying assumptions whether a probiotic has the desired effects in the gut microbial community, we need more studies examining the ecology of probiotic interventions (Walter et al., 2018; van den Berg et al., 2022). A mechanistic, personalized approach to probiotic design—one rooted in empirical metabolic data and ecologic principles—has the potential to propel the field forward.

Previous work trying to predict engraftment has been mostly restricted to fecal microbiome transplant (FMT) studies using non-mechanistic classifiers (Smillie et al., 2018; Podlesny et al., 2021). While potentially predictive, such classifier approaches are sensitive to the underlying assumptions or conditions in which the study is carried out. Because such models are built using statistical methods on data that is assumed to be uniformly collected, these models cannot be generalized to new circumstances. It would be unexpected to see predictions built from patients with diarrhea, undergoing bowel prep, taking antibiotics, and given an FMT accurately predict what happens in patients taking orally administered probiotics. For these reasons, more detailed mechanistic approaches, such as genome-scale metabolic modeling, have been recently explored as more generalizable alternatives (Heinken et al., 2021a; Billard et al., 2021; Jenior et al., 2021).

The goal of this work is to examine the use of metabolic modeling-informed population dynamic approaches. This builds on top of related work in both constraint-based metabolic modeling and population models such as Lotka-Volterra. It is worth highlighting that the use of dynamic flux balance analysis for population models has also been a well-published approach. Despite these successes, there are a number of practical drawbacks for communities of high complexity such as labor-intensive interpretation and high computational complexity.

Population models such as the generalized Lotka-Volterra model are popular tools for understanding microbial community dynamics in a mechanistic manner (Stein et al., 2013; Friedman et al., 2017; Angulo et al., 2019; Kuntal et al., 2019). However, these models are in general difficult to parameterize, with state of the art gradient-matching procedures requiring somewhat dense time-longitudinal data with many replicates (Bucci et al., 2016). Furthermore, it has previously been shown that parameters fit from data to these models do not extend to novel environmental situations, and may even change with the addition of a new taxa to the community (Brunner and Chia, 2019; Momeni et al., 2017). These drawbacks make such mechanistic population models impractical for predicting engraftment. However, by leveraging metabolic modeling we are able to parameterize population models in a way that can be easily adapted to novel environments and does not require dense time-longitudinal data. This allows us to use these models to predict microbial engraftment into a community. See fig. 1 for a comparison of our method with standard parameter fitting.

Genome-scale metabolic models (GSMs) encode the metabolic pathways available in a cell, allowing simulation of cellular metabolism and growth (Lewis et al., 2012). These tools have been used extensively to understand and engineer microbial mono-cultures, and have recently begun to be used to understand and predict the composition and metabolic function of microbial communities (Chan et al., 2017; Diener et al., 2020; Kim et al., 2022). Community methods using GSMs often suffer from a number of drawbacks. These can include a focus on equilibrium states, high complexity and computational cost, and the difficulty of providing accurate GSMs for each community member. By layering simple, pairwise GSM community modeling with population models, we allow for prediction that includes transient behavior (e.g. how quickly an extinction happens within a community) and reduces computational complexity. The availability of accurate GSMs remains a major difficulty, but we demonstrate that progress can be made using algorithmically generated models from whole genomes.

In this paper, we present a method to predict engraftment of an invader into a microbial community in the following manner. First, we construct an interaction network of the microbial taxa found in a sample from the community using pairwise flux balance analysis (FBA) with resource allocation constraints (Kim et al., 2022), which requires a GSM for each taxa. To construct the necessary GSMs, we matched taxa found in the sample to the genomes found in the RefSeq database.
Pairwise GEMs  
Traditional Model Fitting

| Parameter Fitting Calculation | $N^2$ pairwise joint flux balance analysis (linear optimizations) | Gradient matching (quadratic optimization or Bayesian fitting) |
| Free Parameters | 1 joint flux balance analysis meta-parameter + nutrient condition + choice of dynamics | $N^2 - N$ interactions + choice of dynamics |
| Data Required | Single sample compositional data | Dense time-longitudinal data with multiple replicates |
| Generalizability | Can be adapted to nutrient condition | Must refit for each experiment |
| Inherent Assumptions | Microbial optimization for growth, constant direct interactions between microbes | Constant direct interactions between microbes |
| Convenient Assumptions | Uniform intrinsic growth across the community | Constraints on possible interactions (e.g. self-regulation must be negative) |
| Major Practical Limitations | Availability of high quality genome-scale reconstructions and media information | Availability of time-longitudinal data |

**Figure 1.** Population models such as Lotka-Volterra generally require dense time-longitudinal data to accurately parameterize, and directly fit interactions between species (solid arrow in right panel). In this work, we leverage genome-scale metabolic modeling to parameterize population models with only genomic data from a single time-point. This is accomplished by modeling microbial interactions with their shared environment (colored circles in the left panel). (O’Leary et al., 2016), and used these genomes to build models using ModelSEED (Seaver et al., 2021). A complete list of the genomes used can be found in Supplementary File 1. We then simulated pairwise growth using community FBA with growth media adapted from the AGORA genome-scale modeling project (Heinken et al., 2023); precise details are available in S1 Data.

After network construction, we make a prediction based off of simulations with the generalized Lotka-Volterra model (Edelstein-Keshet, 2005; Stein et al., 2013; Friedman et al., 2017), and test the model’s predictive potential by predicting the outcomes of microbiome invasion experiments (Maldonado-Gómez et al., 2016) from the initial presence/absence of species in each sample. The LV model is widely used, but we found that this model could lead to uncontrolled simulated growth. We therefore also investigated alterations to the LV model that dampen the numerical instability.

Despite the drawbacks to both population models and GSMS, limitations in our ability to construct high-quality GSMS, and the limited data requirement of our method, our method produces positive predictive value in the scenario in which it was tested. Our method therefore represents a “low budget” tool that nevertheless can be useful. This success is also evidence that more complex methods that expand on our simple approach could provide great predictive value with a still limited data requirement. Furthermore, we perform two types of sensitivity analysis to demonstrate that the mechanistic nature of the model provides additional insight into the impact of the various components of the network. We perform simulated knock-out experiments to test how sensitive engraftment is to each community member, and we use parameter sensitivity analysis to test how sensitive engraftment is to each network connection.

**Results**

**Predictive Value of the Method**

We first examine the ability of our metabolic modeling-based approach to successfully predict engraftment versus non-engraftment for different microbial species introduced orally across differ-
Table 1. Area under the receiver operating characteristic curves for our method’s predictions of 22 samples from each of two time points (T.P.) using 6 sets of parameters inferred from joint FBA with 6 different sets of hyperparameters. The first three sets of inferred parameters differ in the “resource allocation constraint (RAC)” in joint FBA. We used values of 35 and 70 for this parameter, as well as using joint FBA without RAC. The next three sets of parameters were inferred using an RAC value of 35 but changes to the model environments. EU Average Diet (C Halved/Doubled) had the major carbon sources of the “EU Average diet” (D-Maltose, Sucrose, D-Fructose, and D-Glucose) halved or doubled in availability, and “Complete Medium” simulated the availability of any exchangeable metabolite at uniform simulated inflow.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Baseline T.P. (p-value)</th>
<th>Treatment T.P. (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU Average Diet (RAC 35)</td>
<td>0.6161 (0.1020)</td>
<td>0.8482 (&lt;0.001)</td>
</tr>
<tr>
<td>EU Average Diet (No RAC)</td>
<td>0.6161 (0.1020)</td>
<td>0.8571 (&lt;0.001)</td>
</tr>
<tr>
<td>EU Average Diet (RAC 70)</td>
<td>0.6429 (0.0741)</td>
<td>0.8482 (&lt;0.001)</td>
</tr>
<tr>
<td>EU Average Diet (C Halved)</td>
<td>0.6071 (0.1107)</td>
<td>0.8393 (&lt;0.001)</td>
</tr>
<tr>
<td>EU Average Diet (C Doubled)</td>
<td>0.6339 (0.0808)</td>
<td>0.8304 (0.0010)</td>
</tr>
<tr>
<td>Complete Medium</td>
<td>0.6071 (0.1155)</td>
<td>0.7143 (0.0221)</td>
</tr>
</tbody>
</table>

Our method showed moderate positive predictive value on predictions from the baseline samples, and good predictive value on predictions made from samples taken during treatment. Improved prediction between baseling and treatment samples suggest that a change in microbiome as a response to the introduction of the probiotic impacts our method. This in turn suggests that our method captures at least part of the underlying biological processes that determine engraftment. Furthermore, these results were mostly robust to choice of hyperparameters, with the exception being that the complete medium showed significantly worse predictive value on the samples taken during treatment.

In fig. 2, we show the estimated significance of our results when compared to the null model (720 samples of the null model). We use an approximate “edge swapping” procedure as a null model, which draws interaction parameters randomly from the set of inferred interaction parameters, with replacement. In other words, our null model shuffles the original set of interaction parameters so that any $\beta_{ij}$ in the null model is some $\beta_{ij}$ in the inferred parameter set. This ensures that the network of interactions in our null model has the same edge weight distribution as that of our inferred model so that we may test the significance of the interactions inferred, rather than summary statistics such as diversity.

We also compared our method with two standard machine-learning techniques, a support vector machine (SVM) and a random forest classifier (RF). Unlike our method, which does not use known classifications to “learn” a model, both of these techniques require a “training” set of data.
The power to predict engraftment versus non-engraftment of the *B. longum* probiotic was relatively robust to the joint FBA hyperparameter set-up, as shown measured by the area under the receiver operator characteristic curve (AUC-ROC) (with 1 being perfect classification of engraftment and 0.5 being random). The null model had generally less predictive power than inferred parameters, in particular when predictions were made with samples from the “treatment” time point. The overall improvement in predictions for the “treatment” time point vs. the “baseline” time point suggest that early changes to the microbiome after introduction of a probiotic play a significant role in determining eventual engraftment. P-values were estimated based on over 1000 samples from a null-model.

in which samples are labeled according to their known classification. We trained these models on 16 of the 22 samples (~70% of the data), and tested their ability to predict the remaining 6 samples. We repeated this procedure 1000 times with randomly chosen testing/training sets. For both classifiers, we used the relative abundance data of the set of taxa that our model considered, as well as binarized presence/absence version of the data. We did this because our method only considers presence or absence of taxa, rather than relative abundance. The support vector machine classifier did not show positive predictive value on the data, while the random forest classifier performed, on average across train/test splits, about as well as our method. Our method, therefore, provides the same ability to classify unknown samples as a random forest classifier without the need for any a priori known sample classifications.

**Uniform Shifts in Interaction Parameters**

One consequence of using the generalized Lotka-Volterra model in our method is the possibility of finite-time blow-up in simulation. This happens when one or more taxa approach infinite biomass in finite simulation time. In practice, whenever one or more taxa increases to a level much higher than the initial total biomass of the simulated system, simulation is slowed down and results may be less reliable. There are two simple methods to prevent this from happening. The first is to alter every interaction parameter to be more negative (see eq. (3)), making the system uniformly more antagonistic and less likely to show finite-time blow-up. The second is to include negative self-inhibition (e.g. add $\beta_i < 0$ parameters (see eq. (4)).

We tested both of these alterations by implementing them and computing predictions and the AUC-ROC of those predictions. Figure 4 shows that these changes did not have a significant impact on the performance of our predictions, as long as only antagonism was considered. When we shifted parameters in the positive direction (simulating commensalism or mutualism), finite-time blow-up was more likely and our results were much less stable.
Figure 3. (A) Our method (horizontal lines) significantly outperformed the support vector machine classifier, which was assessed with 1000 random train/test splits. (B) The random forest classifier, also assessed with 1000 train/test splits performed similarly on average to our method.
Figure 4. (A,B) We altered the generalized Lotka-Volterra model with uniform shifts in parameters which either added antagonism or self inhibition to the model. We tested tested self inhibition with values from 0 to 1 (no self promotion) and antagonism with values from 1 (complete antagonism) to −1 (complete commensalism), with all $\beta_{ij} \in [0, 1]$. Neither change had a significant impact on the AUC-ROC of our method's predictions, although adding commensalism to the model made the model and resulting predictions much less stable due to increased finite-time blow-up.

Interpreting Mechanism - Sensitivity to Inferred Parameters

One major advantage to any mechanistic model is that we may measure the sensitivity of our results to perturbations in the model. Here, we investigate the results of perturbing the model of the \textit{B. longum} experiments in two ways - simulating “knock-out” experiments and computing sensitivity to changes in interaction strength.

First, we simulate knock-out experiments by removing a taxa from the full network of interactions. As a result, this organism will be removed from any sample it was previously present in. We then measure the effect this has on our prediction of \textit{B. longum}'s likelihood of engrafting based on the Lotka-Volterra dynamics.

Next, we measure the sensitivity of \textit{B. longum} growth to each interaction parameter in the Lotka-Volterra model. That is, we measure the effect of perturbing each parameter individually on the simulated abundance of \textit{B. longum} at equilibrium according to antagonistic Lotka-Volterra dynamics.

Sensitivity to Community Members

In order to investigate how sensitive \textit{B. longum} engraftment is to each of the other species present in any sample of the \textit{B. longum} experimental data set, we simulate “knock-out” experiments and observe the effect this has on our prediction of engraftment. For each simulated knock-out, we removed a taxa from every sample it was present in and repeated our predictive procedure. We recorded the difference in sample score for engraftment of \textit{B. longum} (eq. (2)) as well as the change in AUC-ROC of our set of predictions. We simulated knock-outs for the five most abundant taxa in the data (averaging across samples).

Table 2 shows the summary statistics of the simulated knock-out experiments. Of the five knock-outs and two sample time-points that we tested, all but one increased predicted engraftment by a slight amount. Interestingly, most of these also increased prediction accuracy slightly, suggesting that the increase in engraftment was more pronounced among the true engrafter samples.

This experiment demonstrates that incorporating mechanism into prediction can provide useful insight beyond prediction. Because our model considers the network of interactions between
Table 2. We experimented with simulated knock-outs for the top 5 taxa in average abundance in the data. The “Sample Proportion” column gives the proportion of samples in the data-set that contain the organism that was knocked out. The Average Score Difference is the average effect of the knockout on our computed engraftment score, with a positive number indicating an average increase in engraftment after knockout of the organism (implying a negative interaction between the organism and \(B. \) longum). The final column shows the impact on our predictions of removing the organism from the analysis.

<table>
<thead>
<tr>
<th>Sample Proportion</th>
<th>Avg Score Difference</th>
<th>AUC-ROC Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline T.P.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Bifidobacterium) adolescentis</td>
<td>0.954545</td>
<td>0.010114</td>
</tr>
<tr>
<td>uncultured (Ruminococcus) sp.</td>
<td>1.000000</td>
<td>0.012803</td>
</tr>
<tr>
<td>uncultured (Clostridium) sp.</td>
<td>1.000000</td>
<td>0.006259</td>
</tr>
<tr>
<td>(Eubacterium) rectale</td>
<td>1.000000</td>
<td>0.006183</td>
</tr>
<tr>
<td>(Faecalibacterium) prausnitzii</td>
<td>1.000000</td>
<td>-0.002250</td>
</tr>
<tr>
<td><strong>Treatment T.P.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Bifidobacterium) adolescentis</td>
<td>0.954545</td>
<td>0.015415</td>
</tr>
<tr>
<td>uncultured (Ruminococcus) sp.</td>
<td>1.000000</td>
<td>0.016707</td>
</tr>
<tr>
<td>uncultured (Clostridium) sp.</td>
<td>1.000000</td>
<td>0.014424</td>
</tr>
<tr>
<td>(Eubacterium) rectale</td>
<td>1.000000</td>
<td>0.011133</td>
</tr>
<tr>
<td>(Faecalibacterium) prausnitzii</td>
<td>0.954545</td>
<td>0.013323</td>
</tr>
</tbody>
</table>

Microbial taxa, we are able to experiment with the effects of each individual taxa by using simulated experiments like knock-outs.

**Sensitivity to Interactions**

Our method is based on the network of interactions between microbial species, which ties together the individual interactions between each pair of species into one complex picture. This means that the interaction between two species unrelated to \(B. \) longum may have an effect \(B. \) longum’s growth. We can compute this effect by computing how \(B. \) longum’s simulated abundance changes as we vary each parameter. Precisely, we can compute the derivative of \(B. \) longum with respect to each parameter \(\beta_{lk}\) in the model.

We computed the sensitivity of our engraftment score to 8 of the interaction parameters in the model, using an RAC value of 35 and the “EU average diet”. For each of the baseline and treatment time-point sample sets, we chose the 2 strongest negative and 2 strongest positive interactions that appeared in at least half of all samples, as well as the 2 strongest positive and 2 strongest negative interactions that have \(B. \) longum as a target (and so act directly on the probiotic) that appeared in at least half of all samples. We observed that engraftment score was very sensitive to all the parameters chosen. In table 3, we show the average sensitivity across the 22 baseline and 22 treatment samples and 8 edges, as well as the average (across edges) variance across the samples. We also include the average (across samples) of the variance in engraftment score across the hyper-parameter choices that we tested (with different environmental conditions and RAC values) for comparison. We notice that, in contrast to the model interaction parameters \(\beta_{lk}\), engraftment score was not particularly sensitive to hyper-parameter choices.

**Discussion**

Our work examines the scenario where the goal of probiotic intervention is a long-lasting alteration of the resident host microbiota. Given that the links between microbes and health have a basis in observations of the resident human microbiota (Gupta et al., 2020), one common assumption is that alterations to resident microbiota would impact health outcomes. Along those lines, we demonstrate that genome-scale metabolic modeling can be used in a simple way to predict the
Table 3. The average sensitivity of engraftment score across 8 parameters and the 22 baseline and 22 treatment samples, as well as the average (across samples) variance across the 8 edges. The 8 edges were chosen because they were the 2 strongest positive edges, 2 strongest negative edges, the 2 strongest positive direct edges (i.e. with *B. longum* as a target) and the 2 strongest negative direct edges. Detailed sensitivity results for these 8 edges can be found in Supplementary File 1.

<table>
<thead>
<tr>
<th></th>
<th>Baseline Time-Point</th>
<th>Treatment Time-Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance Across Set-Ups</td>
<td>4.270608e-06</td>
<td>2.621430e-05</td>
</tr>
<tr>
<td>Average Sensitivity (8 Tested Edges)</td>
<td>3.435107e+33</td>
<td>7.504735e+10</td>
</tr>
<tr>
<td>Variance of Sensitivity (8 Tested Edges)</td>
<td>6.250203e+68</td>
<td>1.626682e+23</td>
</tr>
</tbody>
</table>

outcome of species invasion experiments with minimal study data required. This suggests that

genome-scale models (GSMs) can provide value in understanding microbial community dynamics
from cross-sectional microbiome relative abundance profiling.

Genome-scale metabolic modeling uses genomic data to predict the growth and resource use of
a microbial population by considering the entire internal metabolism of the species. This technique
can be extended to community modeling in a number of ways (Zomorrodi and Maranas, 2012;
Chan et al., 2017; Diener et al., 2020; Kim et al., 2022; Frioux et al., 2020), all with relatively high
levels of complexity. Community dynamics can be inferred from GSMs using a model known as
dynamic flux balance analysis (FBA), which uses a GSM for each taxa in the community to infer
growth rates and dynamic resource usage (Brunner and Chia, 2020). While dynamic FBA provides
a complete picture of community dynamics according to GSMs, it represents a very complex model
that is difficult to analyze and simulate, and still contains a set of unknown parameters.

As a limitation, it is worth noting that all GSM-based methods, including our own, are limited to a
large extent by the accuracy of the metabolic model reconstructions. Low-quality reconstructions
often omit key reactions, or worse, may include reactions that should not have been included.
In all cases, this creates the potential for reaction fluxes or interactions that do not reflect reality.
Improving the quality of metabolic model reconstruction is the subject of challenging and nuanced
research that includes automated pipelines (Benedict et al., 2014; Seaver et al., 2021; Faria et al.,
2023) and hand-curation efforts (Heinken et al., 2023). These efforts are essential to improve the
accuracy of community metabolic modeling, but are beyond the scope of this particular study.

Here, we present a simpler model—a network of emergent interactions between microbes,
whose interaction parameters we can approximate from pairwise genome-scale modeling. The
generalized Lotka-Volterra model and other similar microbial network models are popular tools
for understanding community dynamics (Friedman et al., 2017; Fisher and Mehta, 2014; Angulo
et al., 2019; Bucci and Xavier, 2014). Such models are difficult to parameterize for a myriad of rea-
sons. Good parameterization requires relatively dense time-longitudinal data of absolute, rather
than relative, community abundance (Bucci et al., 2016; Kuntal et al., 2019). By taking advantage
of genome-scale metabolic models (GSMs), we provide a parameterized network that does not re-
quire any time-longitudinal data. In fact, our method as presented here makes use of previously
published GSMs or genomes, meaning that only binary presence/absence information is neces-
sary for a prediction. Additionally, GSMs can be built directly from the genomes found in a sample
using automated tools such as CarveMe (Machado et al., 2018) or ModelSEED (Seaver et al., 2021),
meaning that our method can be extended to include taxa not found in any database.

Another strength of our approach is its interpretability. The *B. longum* engraftment analysis
provides us with an example motivation, which would be to identify other partner microbes that
could improve the responsiveness to the engraftment of a target probiotic. In this instance, we
demonstrate how this might be done by quantifying the impact of five high-abundance microbes
on *B. longum* engraftment. We see, for example, that *B. adolescentis* has a negative impact on *B. longum* engraftment (simulated knockouts increased engraftment score), while *F. prausnitzii* has a
positive impact on *B. longum* engraftment. Although time consuming, a complete set of simulated knockout experiments is readily possible using our method, which could identify the largest positive and negative relationships specific taxa and *B. longum* engraftment.

Inspecting the sensitivity of our method’s engraftment score revealed that it is very sensitive to the individual interaction parameters $\beta_{ik}$ but quite robust to changes in hyper-parameters. This suggests that the method is sensitive to the interactions between microbes, while robust to changes in computational and environmental parameters. It also suggests that individual interactions between microorganisms can have an outsized impact on the downstream composition of a microbial community. This is consistent with the idea that probiotics may have a role in treating human disease by reorganizing the microbiome through the addition of one or a few species.

Lastly, we note that we have previously shown that species-species interaction modeling, i.e. models built from interactions between microbes, do not capture the complexities of microbial community dynamics that emerge as communities change in composition (*Brunner and Chia, 2019*). Here, we mitigate this shortcoming by determining interaction parameters from pairwise models under specific metabolic conditions, providing limited environmental context to our method. However, it is unlikely that these interactions remain constant as the microbial community manipulates its environment. We conjecture that prediction can be improved by accounting for changes in microbial interaction as the environment changes. In upcoming work (*Brunner et al., 2023*), we demonstrate that dynamic FBA implies that a microbial community behaves according to a discrete sequence of interaction networks over time. Incorporating this dynamic behavior may improve prediction with only a modest increase in model complexity, as long as this sequence of networks can be efficiently determined.

### Methods

**Assessing Receptivity with the Generalized Lotka-Volterra Model**

In order to assess how receptive a sub-graph is to an invading taxa, we used the generalized Lotka-Volterra model, which can be associated with any pairwise graph. We simulated the community represented by each sample to equilibrium and scored the performance of the invading taxa by simulated final abundance or time to extinction.

The generalized Lotka-Volterra (gLV) model (*Edelstein-Keshet, 2005*) of a community of $N$ species is written as follows:

$$\frac{dx_i}{dt} = x_i \left(1 + \sum_{j \neq i} \beta_{ji} x_j \right).$$

Notice that we use a version of this model that assumes an intrinsic growth rate of 1 for each taxa because fitting accurate intrinsic growth rates would require additional data, while we wish to assess the performance of our methods without the need for parameter fitting. Our predictions were based on the equilibrium relative abundance of the target variable (e.g. the variable representing a candidate probiotic) in the dynamical system.

The gLV model can display an array of behaviors, including multi-stability and chaos. For this reason, we use a Monte-Carlo sampling approach to generating predictions, taking repeated random draws of initial conditions and simulating forward. In our experiments, we used 1000 draws (i.e. we compute 1000 separate ODE sol) for each sample. In each trial, we scored the network based on the simulated relative abundance of the invader at some large time $T$. It is possible that the invading taxa either dominates the community (i.e. relative abundance approaches 1) or becomes extinct in our simulation. We take this into account by setting a score as follows:

$$s_i = \frac{1}{3} \left( \frac{\text{time to extinction}}{T} + \text{relative abundance at time } T + \frac{\text{time to community dominance}}{T} \right)$$

for each trial $i$ and averaging over all trials.

To test changes in the structure of the generalized Lotka-Volterra model and control for possible finite-time blow-up of model solutions, we tested two general alterations to the model. The first is
a constant shift of the interaction parameters:

\[ \frac{dx_i}{dt} = x_i \left( 1 + \sum_{j=1}^{N}(\beta_{ji} - c)x_j \right) \]  

(3)

and the second is a shift only in the self-interaction terms:

\[ \frac{dx_i}{dt} = x_i \left( 1 - cx_i + \sum_{j \neq i} \beta_{ji}x_j \right). \]  

(4)

For both alterations, we used a single constant \( c \) across the entire model, and experimented with changes in the value of \( c \).

### Inferring interaction parameters

To infer the interaction parameters \( \beta_{ji} \) of the dynamical system, we use a network of interactions implied by a technique known as flux balance analysis (FBA). FBA is a technique used to predict growth rates of microbes using genome-scale information about their internal metabolisms (Lewis et al., 2012). This technique requires a genome-scale metabolic model (GSM) which represents the set of metabolic pathways of a microbe, as well as information about the environmental metabolites available. An optimization problem is then solved to predict a growth rate of the microbe.

Similarly, a technique known as “joint flux balance analysis” (Zomorrodi and Maranas, 2012; Chan et al., 2017; Kim et al., 2022) can be used to estimate growth rates of pairs of microbes using a GEM from each microbe.

Flux balance analysis and joint flux balance analysis allow us to simulate singleton and pair growth experiments in-silico and use the results of these simulated experiments to parameterize our dynamical model. Simulated biomass when grown in a pair was compared to simulated biomass when grown alone in order to determine an implied interaction between microbes. This interaction is then used as a parameter in the dynamical model, which can be conceptualized as an edge in a network of interactions.

To compute growth simulations, we use joint FBA with a set of resource allocation constraints, which have been shown to improve predictions in pairwise experiments (Kim et al., 2022). Briefly, joint FBA with resource allocation constraints solves a linear program defined by the GSMs of the community members. The linear program maximizes total community biomass with the standard FBA constraints as well as the added constraints that each microbe grows a fixed known rate (simulating chemostatic equilibrium), that mass is balanced community wide (with inflow and outflow of nutrients), that there is forced leak of nutrients out of the community (introduced by Kim et al. (Kim et al., 2022) to prevent unrealistically efficient cross-feeding) and that each community member’s total internal reaction flux is bounded (the resource-allocation constraint). Our problem set-up is identical to that described in Kim et al. (Kim et al., 2022), and we provide python code to carry out the computation in our supplemental repository.

To perform joint FBA, simulated inflow of nutrients was determined by medium defined as “EU average diet” by the Virtual Metabolic Human project (Noronha et al., 2019; Ernährungsbericht, 2012), with inflow of 10 mmol/day for nutrients not included in the medium but essential for simulated growth. Growth rates of microorganisms were set to be 0.04 grams/day, and forced leak of nutrients was set to be 0.1 mmol/day. For our main experiment, we set the RAC constraint to 35 mmol/day as was shown to lead to accurate prediction in Kim et al. We also adjusted this value, as well as the defined medium, in order to assess our method’s robustness to meta-parameters.

In order to test the impact of joint FBA hyperparameters including medium and resource allocation on our predictions, we also created networks of interaction parameters by performing joint FBA with five alternative sets of hyperparameters. Two of these networks were made with alterations to the RAC constraint we used in computing joint FBA. We computed interactions using an RAC value of 70, which relaxes the constraint on total flux, and without a resource allocation constraint. The final three networks were constructed with different media, simulating differences
in the external environment. We constructed a network with the major carbon sources of the “EU Average diet” halved in availability, and a second with these carbons doubled. The major carbon sources were taken to be D-Maltose, Sucrose, D-Fructose, and D-Glucose. Finally, we computed a network with a “complete medium” in which any exchangable metabolite for each model was made available with uniform simulated inflow.

We use the log-ratio of simulated growth in pairs and alone as the implied effect of one microbe on another. That is, if simulated growth of species $i$ alone is $x_i$, and growth of species $i$ when coupled in a pair with species $j$ is $x_{ij}$, we weight the edge from species $j$ to species $i$ as

$$w_{ji} = \log \left( \frac{x_{ij}}{x_i} \right)$$

We use the log-ratio because it is the simplest choice of translating simulated data to interaction that accounts for fold-changes. Of course, the simple difference in simulated biomass is another possible choice, but this will be much more sensitive to absolute scaling in simulation, so, for example, slow growing taxa will be biased towards smaller interactions in the network. Furthermore, our in-silico growth experiments only provide simulated biomass values at chemostatic equilibrium, meaning that fitting Lotka-Volterra parameters directly to these data is an underdetermined problem, particularly when pair growth leads to a simulated extinction.

We also re-scale our network so that all edge weights are in the interval $[-1, 1]$, meaning we take as our edge weights

$$w_{ji} = \frac{w_{ji}^*}{\max_{k,l}(|w_{kl}^*|)}.$$  

This re-scaling is done for computational convenience, and can be viewed as a time re-scaling of the network dynamics which does not effect our predictions. Additionally, we tested the approach after adjusting either all parameters by addition of a single constant value to every $\beta_{ji}$, as well as by adjusting the parameters by addition of a single constant value to every self-regulatory parameter $\beta_{jj}$.

Creation of GSMs
In order to carry out our analysis, we needed to construct GSMs for the taxa found in the data. To do this, we matched taxa in the data by NCBI taxa ID to genomes in the NCBI refseq database(O’Leary et al., 2016). We matched taxa to the nearest (by tree distance according to NCBI’s taxonomic database) genome that was listed as “Chromosome” or “Complete Genome”, or contained at least 1000 genes. On average, this approach allowed us to perform our experiments with $\sim 87 - 88\%$ coverage of each sample. Taxa for which we failed to match a genome were removed from the data, and all analysis was performed on the remaining. For full details on the coverage of each sample, see Supplementary File 1.

Next, we used the ModelSEED database through ModelSEEDPy(Seaver et al., 2021) to reconstruct GSMs for each genome that we matched to the data. All models constructed showed positive simulated growth rates using FBA on the “EU average diet” medium.

Prediction and evaluation
We used data from Moldonado-Gómez et al.(Maldonado-Gómez et al., 2016) in order to test the predictive power of the method. This data consisted of bacterial community composition of fecal samples taken over the course of experiments in which Bifidobacterium longum AH1206 was administered as an oral probiotic to 23 subjects, with data available for 22 of these. Subjects were then differentiated into "engrafters" and "non-engrafters" based on the survival of the probiotic strain, as determined by cell culture at later time-points, allowing us to use our method as a classifier. We note that our method "learns" using joint FBA, and so there is no need to split the data into training and testing sets. In fact, the "learning" procedure does not make use of any known sample classifications, as a traditional machine-learning algorithm would. This allows us to test our
Figure 5. Schematic of the modeling process. In brief, we generate an interaction network of genome-scale models using pairwise joint flux balance analysis. To produce a prediction of engraftment for a given sample, we use the taxa present in the sample to generate an induced sub-graph of the full network. This is then used to define the parameters in the generalized Lotka-Volterra dynamical system to generate a prediction of engraftment.
ment” sets of samples.

### Sensitivity to Community Members and Interaction Parameters

We compute *B. longum* AH1206 engraftment predictions using simulated knock-out experiments simply by removing select taxa from the model. The result is a prediction score computed according to eq. (2). This score can be compared to the non-knockout experiment score to determine if the knock-out increased or decreased the predicted probability of *B. longum* engraftment. We compute and report the relative change in predicted engraftment score as

\[
\Delta = \frac{(\text{score with knock-out taxa removed}) - (\text{score with knock-out taxa included})}{(\text{score with knock-out taxa included})}.
\]

We compute the sensitivity of *B. longum*’s growth to an interaction parameter $\beta_{kl}$ directly by using the chain rule (see, for example ([Zi, 2011](#)). This has the following form:

\[
\frac{\partial}{\partial t} \left( \frac{\partial x_i}{\partial \beta_{kl}} \right) = x_i x_i \delta_{i,j} + x_i \sum_{j=1}^{n} \beta_{ij} \frac{\partial x_j}{\partial \beta_{kl}} + \left( 1 + 2 \beta_{ii} x_i + \sum_{j=1}^{n} \beta_{ij} x_j \right) \frac{\partial x_i}{\partial \beta_{kl}}.
\]

Equation (8) allows us to solve a system of differential equations to determine the sensitivity of invader growth to that parameter. We solve this system of equations to a large time $T = 100$ and report the weighted average of some late-time interval $[t, T]$, with later time-points weighted more heavily, using the final 40 simulation time-points.

### Acknowledgements

The authors would like to thank Dr. Marie Kroeger for her thoughtful comments in helping to prepare this article. J.B. was supported in this work by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division using Award number F255LANL2018 and the Los Alamos National Laboratory Center for Nonlinear Studies. This study was partially supported by funds from Mayo Clinic’s Center for Individualized Medicine.

### Data & Code Availability

Data from the study that we used to evaluate our method can be found at the following source:

- ([Maldonado-Gómez et al., 2016](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA324129/))

Our method is implemented as the `friendlyNets` package, available for download at [https://github.com/lanl/friendlyNets](https://github.com/lanl/friendlyNets) along with re-formatted data and python scripts for the analysis found in this paper.

### References


Brunner JD, Chia N. Minimizing the number of optimizations for efficient community dynamic flux balance analysis. PLoS computational biology. 2020; 16(9):e1007786.


Supplementary File 1 (S1_Data.xlsx) contains the following tables:

- Relative abundance of each taxa in each sample in the data-set, product of Bracken analysis on original data
- Genomes matched to taxa in data from RefSeq database, used to create models
- Names of taxa in the data, matched with Taxa ID
- Coverage (as proportion of relative abundance) of models used in analysis of Baseline time-point samples (i.e. taxa for which we could identify a high quality close match genome)
- Coverage (as proportion of relative abundance) of models used in analysis of Treatment time-point samples (i.e. taxa for which we could identify a high quality close match genome)
- Main media file used, from vmh.life
- Cell counts of B. longum probiotic, provided by Maldonado-Gomez et al.
- Summary of parameter sensitivity results (average and variance across samples)
- Parameter sensitivity in baseline time-point samples (2 strongest positive and 2 strongest negative edges, 2 strongest positive and 2 strongest negative edges with B. longum as target)
- Parameter sensitivity in treatment time-point samples (2 strongest positive and 2 strongest negative edges, 2 strongest positive and 2 strongest negative edges with B. longum as target)