Metabolite mediated modeling of microbial community dynamics captures emergent behavior more effectively than species-species modeling.

J.D. Brunner$^{1,2}$ and N. Chia$^{1,2}$

$^1$Division of Surgical Research, Department of Surgery, Mayo Clinic
$^2$Microbiome Program, Center for Individualized Medicine, Mayo Clinic

Abstract

Personalized models of the gut microbiome are valuable for disease prevention and treatment. For this, one requires a mathematical model that predicts microbial community composition and the emergent behavior of microbial communities. We seek a modeling strategy that can capture emergent behavior when built from sets of universal individual interactions. Our investigation reveals that species-metabolite interaction modeling is better able to capture emergent behavior in community composition dynamics than direct species-species modeling.

Using publicly available data, we examine the ability of species-species models and species-metabolite models to predict trio growth experiments from the outcomes of pair growth experiments. We compare quadratic species-species interaction models and quadratic species-metabolite interaction models, and conclude that only species-metabolite models have the necessary complexity to explain a wide variety of interdependent growth outcomes. We also show that general species-species interaction models cannot match patterns observed in community growth dynamics, whereas species-metabolite models can. We conclude that species-metabolite modeling will be important in the development of accurate, clinically useful models of microbial communities.

1 Introduction

The microbial communities of the human body, collectively called the “human microbiome”, act on the host in a symbiotic relationship which can have profound effects on health and disease [1, 2, 3, 4, 5, 6, 7]. This can be seen in the impact of bacterial colonization on the development of the adaptive immune system [7] as well as the many observed microbiome alternations in diseases ranging from multiple sclerosis [8] to colorectal cancer[4]. Importantly, these changes go beyond the presence or absence of a single species, but derive from more complex shifts in community composition. One prime example of this is the presence of pathogenic bacteria among the microbiota of disease-free asymptomatic individuals [7]. It is clearly not enough to identify and target a single species when attempting to explain the impact of the microbiome on host health. Instead, we must understand the interactions within a microbial community. These interactions determine whether a potentially pathogenic bacteria behaves as a beneficial, neutral, or pathogenic member of the microbial community. Because these properties arise from the broader community, rather than the individual species, we call these emergent properties. Identifying and predicting microbial community composition and the resultant emergent properties are an important part of disease prevention, diagnosis, and treatment in the burgeoning world of data-driven and individualized medicine.

In order to understand the composition of the microbiome, we must understand the dynamic process of growth, invasion, and extinction which leads to a stable microbiome in healthy individuals, as well as the dynamic responses of the microbiome to changes in host health and diet. Such an understanding would give us the fundamental rules by which to alter the microbiome in an effective and stable way. We may then prevent disease and improve disease outcome by formulating these rules into a model of how microbiome composition changes in response to perturbation, and use this model to design treatments to manipulate the microbiome.
The goal of this manuscript is to identify a modeling framework for recapitulating community growth dynamics from sets of fundamental interactions. This modeling framework must be extensible, that is, allow us to directly combine models of smaller communities to create models of composite communities without discovering new parameters. As part of this goal, we would like this model to be “as simple as possible but no simpler”. These properties would allow us to generate a clinically useful dynamical model of the microbiome—one that can predict community composition dynamics from individualized information such as initial composition and environmental perturbations due to treatment. It is worth noting that in this setting, a single patient cannot in general provide enough data to accurately parameterize a purpose-built model. Instead, an individualized model needs to be built using parameters generated from previous experiments or other data-sets.

This contradiction can be overcome by using a modeling framework that infers whole community dynamics from simple, fundamental interactions that are assumed to be universal. Universal building block interactions could be used to build purpose-built predictive dynamic models in an “n of one” manner, meaning built with data from a single patient [9, 10]. It is as yet an open question as to the nature of these building blocks, and indeed if any can be found [11, 12, 13, 14, 15]. Here, we examine two popular modeling frameworks, species-species interaction (SSI) and species-metabolite interaction (SMI) models, and evaluate them using interdependent growth experiments of single species, pairs, and trios from Friedman et al. [16]. These growth experiments were carried out on flat bottom plates with an experimental growth media and serial dilution. We assume well mixed, spatially homogeneous interactions and growth. We are therefore testing this on the simplest possible situation - pair and trio growth experiments, asking whether or not a modeling strategy can recapitulate the observed outcomes of these experiments.

Perhaps the most popular candidate for a set of fundamental modeling building blocks is the set of interactions between species of microbes [16, 17, 18, 19, 20, 21]. We call such models species-species interaction (SSI) models. This strategy follows from microbial co-occurrence networks, which can be inferred from 16s rRNA gene or metagenomic sequencing data [22, 23]. Focusing on species-species interactions is notably the strategy of the popular Lotka-Volterra (LV) model and its generalizations [16, 17, 18, 19]. The Lotka-Volterra model reproduces the dynamics of interacting species according the law of mass action [24, 25]. Therefore, it is an appropriate model of direct interaction between species in a well mixed and stable environment. The Lotka-Volterra model and SSI models in general may therefore be useful when fit to stable environments.

SSI models can capture some of the emergent behavior of composition dynamics, but fail to capture higher order interactions which require more than two species [26, 27]. We find in general that SSI models imply a strict condition on growth dynamics that is not observed in data. Furthermore, we find that the quadratic SSI model, usually called the “generalized Lotka-Volterra (gLV) model”, is not capable of recapitulating the entire set of pair and trio growth outcomes using a single parameterization. Although the gLV model may be useful when fit to whole communities [19, 20, 21], our work suggests that this model lacks the necessary complexity to be predictive when built from building blocks assumed to be universal.

Alternatively, the microbiome can be modeled by species-metabolite interaction (SMI) models, which are constructed using the interactions of individual microbes with a shared metabolite pool [28, 29, 30, 14, 11]. SMI models follow from networks which include both microbiota and metabolites, which may be inferred from literature [30] or from genome-scale metabolic networks [31]. Recently, Momeni et al. [11] proved that SMI models are strictly more complex than SSI models. Like SSI models, SMI models may include arbitrary complexity in interaction terms; saturating kinetics (e.g. Michaelis-Menten or Hill kinetics) are a particularly common choice [32, 28, 29]. We show that a simpler quadratic species-metabolite interaction (QSMI) model can recapitulate the growth experiment outcomes of Friedman et al. [16] with a single parameterization.

It is worth noting explicitly that our work does not explore the accuracy of specific modeling tools [33, 34, 35, 36], but instead examines whether the mathematical formulation of the model could ever be used to recapitulate the biological dynamics. A positive answer means that the mathematical form of the model can potentially be useful, indicating promise for future endeavors, while a negative answer indicates that the basic formulation of the model is inappropriate for predictive models of microbial communities. More specifically, we investigate this question by inspecting to what extent models of simple communities can be used to build accurate models of larger communities. Precisely, we asked whether these models have the capacity for a parameterization that recapitulates qualitative outcome of both pair and trio growth experiments.
2 Model definitions and background

2.1 Species-Species Interaction (SSI) models

Species-Species Interaction (SSI) models are dynamical models of the composition of a community of organisms \(i, j, k, \text{ etc.}, \) built by assuming direct interaction between species. We model this by assuming that the population size of some species changes as the product of some per-organism growth rate and the current population size. This per-organism growth rate is then determined by interactions with other species.

SSI models are popular in ecology, including in the study of the microbiome, because they are computationally simple to create and analyze \([16, 17, 18, 20, 37, 38, 39]\). They are often fitted to large communities of microbiota, and interactions between species are assumed from this fitting \([19, 21]\). This fitting implies a relationship between species, and these relationships are often classified as competitive, mutual, or parasitic \([40]\). In the study of the human microbiome, discovering interactions between species is an active area of research \([18, 41, 20]\), and automated tools \([18, 42, 43, 44]\) for the construction of SSI models are used \([18, 42, 43, 44]\). It is therefore of interest to understand whether or not these interactions are preserved as the community changes, and so can be used as universal building blocks.

The general form of an SSI model is as follows:

\[
\frac{dx_i}{dt} = x_i \left( f_i(x_i) + \sum_{j \neq i} h_{ij}(x_i, x_j) \right) 
\]

where \(x_i\) represents the biomass of organism \(i\) and the functions \(f_i(x_i)\) and \(h_{ij}(x_i, x_j)\) respectively represent lone growth of organism \(i\) and the effect of organism \(j\) on growth of organism \(i\). Note that we allow complete generality, including the case of distinct functions \(h_{ij}\) for each pair \(i, j\). It is convenient to write SSI models as product of current population size \(x_i\) and a per organism growth rate \(\left( f_i(x_i) + \sum_{j \neq i} h_{ij}(x_i, x_j) \right)\). The number of interaction terms in an SSI model scales with the number of pairs of \(N\) microbes, and so scales as \(N^2 - N\).

In this manuscript, we require that \(h_{ij}(x_i, x_j)\) satisfy \(h_{ij}(x_i, 0) = 0\) and that \(h_{ij}(x_i, x_j)\) not switch sign (i.e. \(h_{ij}(x_i, x_j) \geq 0\) or \(h_{ij}(x_i, x_j) \leq 0\) for any non-negative population sizes \(x_i, x_j\)). This simply means that species \(j\) must be present to have an effect on the growth of \(i\), and that species \(j\) either increases the growth of species \(i\) or decreases this growth (or has no effect), regardless of the population sizes of species \(i\) and species \(j\), while the strength of this effect may depend on population sizes.

The generalized Lotka-Volterra (gLV) model

The simplest SSI model assumes direct, pairwise interactions proportional to species concentration, and is called the generalized Lotka-Volterra (gLV) model. In this model, the per-organism growth rate of a population changes proportionally to the size of each other population. Such changes may be positive or negative, indicating mutualism, parasitism, predation and competition.

The generalized Lotka-Volterra model faithfully models direct pairwise interactions between agents (e.g. physically interacting organisms or reactants in an industrial reactor) under the assumption of mass action kinetics \([45, 25]\), but does not include any possible environmental variation in interaction. It is therefore a fair representation of species-species interaction in a controlled environment. Because of this, the generalized Lotka-Volterra model is well studied and its parameters are often inferred from correlations seen in available 16S rRNA gene or metagenomic sequencing data \([22, 20]\). Furthermore, the generalized Lotka-Volterra model is commonly used to infer relationships between species and model the microbiome \([37, 38, 39, 18, 42, 43, 44]\).

This model is typically written \([27, 16]\) as follows:

\[
\frac{dx_i}{dt} = r_i x_i \left( 1 - \frac{1}{K_i} x_i + \sum_{j \neq i} \alpha_{ij} x_j \right) 
\]

where \(r_i > 0\) is the intrinsic growth rate of the community of organism \(i\), \(K_i > 0\) is the carrying capacity of the environment for the organism, and \(\alpha_{ij}\) is the interaction between species. Although pairwise and linear...
in per organism growth rate, this model can display a wide range of behaviors seen in nature, including invasion, competitive exclusion, coexistence, and multi-stability (i.e., stable long-term outcomes that are dependent on initial community sizes even for fixed parameters) \[46, 27\].

2.2 Species-Metabolite Interaction (SMI) models

Species-metabolite Interaction (SMI) models, also called metabolite (or resource) mediated models, use the interaction of a microbe with an environmentally available metabolite as their fundamental interaction\[28, 14\]. Similar to SSI models, we assume that the population size of some species changes as the product of some per-organism growth rate and the current population size. In SMI models, per-organism growth rate depends on available metabolites rather than a fixed carrying capacity. Additionally, metabolites are used and produced by individual species, in many cases as a by-product of some process involving another metabolite. Interactions between microbes are then possible through manipulation of the shared metabolite pool.

SMI models have recently become of interest in the study of the human microbiome as data on the metabolite pool has become available \[32, 47, 48\]. Techniques for integrating species abundance and metabolomic data are now being developed to understand the mechanisms of microbiota organization \[49, 40\].

We define a SMI model to be any model of the type

\[
\frac{dx_i}{dt} = x_i \sum_{j=1}^{m} f_{ij}(x_i, y_j) \\
\frac{dy_j}{dt} = \sum_{i=1}^{n} \sum_{l=1}^{m} h_{il}^j(x_i, y_l)
\]

where \(n\) is the number of microbial species and \(m\) the number of metabolites modeled, \(x_i\) again represents biomass of organism \(i\), and \(y_j\) is the concentration or biomass of metabolite \(j\). As species and metabolites are added to the model, the number of terms could grow as large as \(NM + NM^2\), where \(N\) is the number of species and \(M\) the number of metabolites included.

The simplest SMI model has competition for resources, leading to the well known competitive exclusion principle which states that an environment cannot support more species than it has food sources. However, metabolite mediated models have more versatility, and do not need to hold to that principle \[29\].

Popular choices of interaction terms \(f_{ij}(x_i, y_j)\) and \(h_{il}^j(x_i, y_l)\) include sigmoidal (i.e. saturating) kinetics, such as Michaelis-Menten or Hill kinetics \[29, 32\], which have a diminishing change in effect as concentrations of metabolite increase, and polynomial kinetics.

The Quadratic Species-Metabolite Interaction (QSMI) model

The simplest SMI model is quadratic, and includes consumption and production of metabolites by microbes. This model assumes that per-organism growth rate of microbiota changes proportionally to the amount of each metabolite that it interacts with.

Similar to the gLV model, the QSMI model is a faithful model of interacting components of a well mixed system \[45, 25\], without allowing for any variation in environmental variables. In order to model stable growing communities, we include constant dilution of microbiota as well as metabolites. The QSMI model is the closest analogue among SMI models to the generalized Lotka-Volterra model, as both are quadratic polynomials that faithfully model well-mixed interacting actors.

This gives the general model for \(n\) species and \(m\) metabolites

\[
\frac{dx_i}{dt} = x_i \left( \sum_{j=1}^{m} \psi_{ij} y_j - d_i \right) \\
\frac{dy_j}{dt} = f_j - d_j^* y_j - y_j \sum_{i=1}^{n} \kappa_{ij} x_i + \sum_{i=1}^{n} \sum_{l=1}^{m} \phi_{il}^j x_i y_l
\]
where a standard competition model would have $k_{ij} = \psi_{ij} \geq 0$. This model assumes that a microbial population’s growth rate depends on the availability of the resources being metabolized for growth, and these resources are depleted as this growth happens. The terms $\phi_{ij}x_iy_l$ allow for the possibility that metabolites are produced as by-product of some microbial metabolic pathway.

3 Results

Recall that our goal is a model of a composite community that can be built by directly combining models of smaller communities without discovering new parameters. We therefore analyze SSI, gLV, SMI, and QSMI models with the goal of demonstrating which of these frameworks has the capacity for such a model. We find that gLV, and even general SSI models, cannot achieve this goal. However, the additional complexity of QSMI models is enough to allow a model which recapitulates the outcomes of pair and trio growth experiments from Friedman et al.[16] with a single parameter set. We conclude that SMI models show the most promise in building models of larger microbial communities.

3.1 Reversal of qualitative effects in general SSI models

One consequence of SSI models, regardless of the specific choices of interaction functions $h_{ij}$, is that they imply a classification of the interaction between two microbes. That is, the sign of $h_{ij}(x_i, x_j)$ indicates if microbe $j$ has a positive effect on microbe $i$ ($h_{ij}(x_i, x_j) > 0$), or a negative effect on microbe $i$ ($h_{ij}(x_i, x_j) < 0$); recall that $h_{ij}(x_i, x_j)$ does not change sign for non-negative $x_i, x_j$. Generally, relationships are classified using both $h_{ij}(x_i, x_j)$ and $h_{ji}(x_j, x_i)$[40]. SSI models allow us in some cases to predict the sign of the combined effects of two microbes on a third. Precisely, if two microbes have the same qualitative effect on the growth of a third, SSI models imply that their combined effect should be qualitatively the same (although of course different in magnitude). For example, if $h_{ij}(x_i, x_j) > 0$ and $h_{ik}(x_i, x_k) > 0$, then clearly $h_{ij}(x_i, x_j) + h_{ik}(x_i, x_k) > 0$ for all non-negative $x_i, x_j, x_k$, and we should observe increased growth in microbe $i$ when in a trio with these two other species as compared to when grown alone.

We estimate the effects on growth of microbes in pairs and trios by using the time-course data from Friedman et al.[16] to directly estimate the per-organism growth rate term of eq. (1). We can therefore compare the effect on the per-organism growth rate of organism $i$ of being grown with species $j$ or species $k$ to the effect of being grown with both $j$ and $k$. We expect that if $j$ and $k$ both increase per-organism growth rate of species $i$, then the combined effect of species $j$ and $k$ is to increase per-organism growth rate of $i$, as in fig. 2(a). However, we do not always see this. In fact, we find that 11 of the 54 trios have at least one reversal of effect. Figure 2(b) shows an example in which two species have a positive effect on the growth of a third when grown in pairs, but when grown in a trio there is reduced growth in the third species. All of the implied pairwise relationships are shown in fig. 1(a). The list of all trios with pair and trio estimated effects can be found in the supplemental file qualitative.csv, and the 11 reversals are also found in reversals.csv.

3.2 Parameter fitting in the gLV model

Above, we show that SSI models in general will not match the dynamics of the growth experiments from Friedman et al.[16]. However, it may be that SSI models, including the popular gLV model, give the correct qualitative outcomes in terms of survival and extinction over long time scales. In order to determine if this is the case, we fit a parameter set (the set of $\alpha_{ij}$ in eq. (2)) to the time-course data of pair growth experiments, and ask if the model’s asymptotic stability correctly predicts the outcome of the trio growth experiments. That is, we take as a model’s “prediction” the result of linear asymptotic stability analysis (see 4.2).

We find that parameters fitted to pair growth experiments explain trio experiments for half of all trios, meaning that the model correctly predicts survival/extinction outcomes for half of all trios. Note that Friedman et al.[16] report that parameters fit to pairs lead to accurate predictions of 84% of trios. However, this was calculated using a different definition of model prediction (see 4.2 for details). Figure 1(b) shows the interactions implied by this parameter fitting. Interestingly, these interactions do not match the network of interactions determined using change in time-averaged growth, shown in fig. 1(a). 5
Figure 1: (a) Relationships between microbial species as implied by estimating average per-organism growth rate from growth data. (b) Relationships between microbial species as implied by parameters of the gLV model fitted to pair growth experiments. (c) Relationships between microbial species as implied by parameters of the gLV model from computational search for parameters to explain all experimental outcomes. (a,b,c) Blue arrows indicate positive influence on growth, red arrows indicate negative influence, and nodes are labeled with the species used, as abbreviated in Friedman et al.\cite{16} (Ea (Enterobacter aerogenes, ATCC#13048), Pa (Pseudomonas aurantiaca, ATCC#33663), Pch (Pseudomonas chlororaphis, ATCC#9446), Pci (Pseudomonas citronellolis, ATCC#13674), Pf (Pseudomonas fluorescens, ATCC#13525), Pp (Pseudomonas putida, ATCC#12633), Pv (Pseudomonas veronii, ATCC#700474) and Sm (Serratia marcescens, ATCC#13880)). (d) Proportion of edges in the network which match qualitatively (i.e. have the same sign) for each pair of networks from (a), (b), and (c).
Figure 2: (a,b) Arrows indicate influence on per-organism growth rate, as determined by deviation from growth rate when grown alone. Combined arrows represent the implied sum of influences. (c,d) Deviation in per-organism growth, estimated from time course data reported in Friedman et al.[16] (dashed line showing no deviation) (a,c) Pa (Pseudomonas aurantiaca, ATCC#33663) and Sm (Serratia marcescens, ATCC#13880) both have positive effects on the growth of Ea (Enterobacter aerogenes, ATCC#13048) when grown in a pair. When these three are grown in a trio, the growth of Ea is promoted relative to its growth when grown alone. This trio makes sense in the context of SSI models. (b,d) Pch (Pseudomonas chlororaphis, ATCC#9446) and Pv (Pseudomonas veronii, ATCC#700474) both have positive effects on the growth of Pp (Pseudomonas putida, ATCC#12633) when grown in a pair. However, when these three are grown in a trio, the growth of Pp is reduced compared to its growth when grown alone. This trio does not make sense in the context of SSI models.
Figure 3: Monte-Carlo experiments of the stochastic generalized Lotka-Volterra model, with parameters fitted to deterministic pair model. Darker color indicates that a larger proportion of the 1000 trials used matched the observed growth experiment outcomes. Pairs and trios are abbreviated as in Friedman et al.[16].

Additionally, we inspect long-time simulations of the gLV model for trios using parameters fitted to pair growth experiments, in order to determine if oscillatory behavior is predicted. We observe no oscillatory behavior in the simulations.

In order to determine if the model’s failure to recapitulate experiments is the result of random fluctuations in growth, we use a stochastic version of the gLV model with the same parameter set. We determine the likelihood of the observed experimental outcome according to the stochastic gLV model, and see in fig. 3 that many outcomes have very low likelihood. Our experiment with the stochastic generalized Lotka-Volterra model shows that random fluctuation in growth is unlikely to explain the failure of the generalized Lotka-Volterra model to match the growth experiment data.

3.3 The gLV model’s capacity to recapitulate experimental outcomes

We next determine if the gLV model has the capacity to recapitulate the outcome of the trio growth experiments while keeping a single parameter set that recapitulates pair experiment outcomes. We define the outcome of trio growth experiments in a binary fashion as reported by Friedman et al.[16], and similarly define the outcome of pair growth experiments based on the final time-point of the experiment.

Finding a single set of $\alpha_{ij}$ such that the model correctly predicts every pair and trio growth experiment would imply that the gLV model has the necessary complexity to match the qualitative outcomes of the growth experiments, if not the dynamics. However, we are unable to find this parameter set using a computational search with a pseudo-genetic algorithm (see 4.4). Indeed, with the parameters resulting from this search, the model correctly predicts only 59% of the trio outcomes. This suggests that the set of parameters for the gLV model which cause it to recapitulate the growth experiments is small or empty.

To reduce the search space, we divide the growth experiments and attempt to find a parameter set for each group that predicts all of the qualitative outcomes of the growth experiments in that group. We use as groups all of the experiments involving one or both of some pair of microbial species (note that these groups are overlapping, but treated independently). For only 9 of the 28 pairs, parameters could be found that explained each trio involving that pair. Figure 4 shows the proportion of trios involving a given pair...
Figure 4: Proportion of explained trios as a result of each pair experiment, using the fitted interaction between that pair and varying the interactions involving the third member of a trio. For some pairs, parameters that explained the data are found for any third microbe, while for some parameters cannot be found for most third microbes. Pairs and trios are abbreviated as in Friedman et al.[16].

Figure 5: Network representation of a cross-feeding chain in which metabolism of $y_1$ by $x_2$ produces a byproduct which $x_1$ can metabolize for growth, and this process in turn produces a byproduct which $x_3$ can metabolize for growth. Note that the production of $y_2$ by $x_2$ only occurs if $y_1$ is present, and the production of $y_3$ by $x_1$ only occurs when $y_2$ is present, so that this chain does not have an effect on other trios involving a subset of three species in the chain. For simplicity, this dependency is not shown, nor is it included in fig. 6 which can be correctly predicted.

3.4 A QSMI model to recapitulate growth experiments

We ask if there exists a quadratic species-metabolite model which can match the qualitative outcomes of the growth experiments presented in Friedman et al.[16]. As with the generalized Lotka-Volterra model, we seek a single parameter set that can explain the interdependent growth experiments. We can build such a model by assuming the existence of one initially present metabolite along with additional molecules produced by the microbiota. These additional molecules allow us to form cross-talk chains, as shown in fig. 5. This model also demonstrates that the situation detailed by fig. 2(b) can be modeled by a QSMI, using the mechanism of fig. 5 with $y_3$ acting instead as a poison to reduce growth of $x_3$.

For the pair experiments of Friedman et al.[16], we need to add 19 cross-feeding molecules, bringing the total metabolites to 20. The pair models lead to trios for which we need to add cross-feeding chains to prevent a single extinction for 4 trios and prevent two extinctions for 1 trio. We also must implement cross-poisoning to cause a single extinction for 19 trios. For 1 trio, we need to adjust the model to cause one extinction and prevent another.

In total, we have a single model of 8 microbes and 72 molecules which are part of part of 19 pair-specific and 25 trio-specific cross-talk pathways. This model, when restricted by initial state to only two or three microbial species, recapitulates the outcomes of the growth experiments.

Figure 6 shows the network of metabolite mediated interactions of the QSMI model. In this model, every microbe grows on a single metabolite, labeled $y_1$, and various pair or trio cross-talk chains alter that growth by providing extra resources (cross-feeding) or inhibiting a microbe’s growth (cross-poisoning).
Figure 6: Schematic of the full QSMI model that recapitulates growth experimental outcomes. This model contains the 8 microbes grown in Friedman et al.[16], as well as 69 metabolites (squares), and 3 signalling molecules (diamonds). Blue arrows from metabolite to microbe indicate that the microbe uses that metabolite to grow, while red arrows from metabolite to microbe indicate that the metabolite inhibits growth of the microbe. Blue arrows from microbe to metabolite indicate production, and red arrows from microbe to metabolite indicate consumption or degradation. To match qualitative outcomes of growth experiments in Friedman et al.[16], the model is run with only Y1 initially present, and initial microbial populations according to the growth experiment.
3.5 Complexity of the QSMI model that recapitulates trio experiments.

We wish to estimate the complexity of a QSMI that explains a given set of growth experiments. To do this, we reduce our set of allowable QSMI models to those that only include direct cross-talk as well as simple cross-talk chains such as that shown in fig. 5, and detailed in eqs. (17) and (18). This restriction allows us to automate construction of a model that explains all but 2 of the trio experiments from Friedman et al.[16] using only 17 metabolites (this network can be viewed in min_met_network.csv in the supplemental repository).

We next estimate the required complexity of a general QSMI model that recapitulates experiments. For a randomly generated set of growth outcomes, we determine how well an automatically generated QSMI model can recapitulate these outcomes, and how complex such a model must be. We hope that the best model generated for a given outcome set is of reasonable accuracy, explaining most or all of the data set, and complexity, requiring not too many metabolites. Figure 7 shows a histogram of the coverage of (i.e. what proportion of a set of outcomes is recapitulated) and number of metabolites needed in the best model we generate for a sample of randomly generated “experimental” outcomes. The coverage achieved follows roughly a normal distribution with mean 0.754 and standard deviation 0.053, while the number of metabolites also follows roughly a normal distribution with mean 26.67 and standard deviation 2.54. This experiment demonstrates that QSMI models have the power to explain the majority of outcomes in a data set with a reasonable number of metabolites, and without pathways more complex than the one shown in fig. 5.
4 Methods

4.1 Growth data used

We use data from growth experiments published in Friedman et al.[16]. These experiments were carried out in flat-bottomed plates and grown in 48 hour growth-dilution cycles. Cell density was assessed using optical density (OD), and relative abundance was assessed by plating and colony counting. For this manuscript, we use units of OD computed by multiplying total culture OD by fraction of each species. Growth data is available in the supplemental material folder friedman_et_al_data. For single microbe and pair growth experiments, Friedman et al.[16] provided intermediate time-course data. However, for trio growth experiments, only beginning and endpoints were reported, and so used here.

4.2 Defining model prediction

We define model prediction by the behavior of the model as time approaches infinity for any positive initial conditions. We compare this model prediction to experimentally observed extinction and coexistence, as reported in Friedman et al.[16].

All models that we consider are ordinary differential equations (ODEs), with the exception of a stochastic analogue to an ODE model. For an ODE model, we define a model’s outcome from the asymptotic stability of equilibrium, using standard methods (see for example [27]). That is, we write any ODE in the form

\[
\frac{dx}{dt} = f(x) \tag{7}
\]

where \( x \) and \( f \) may be vector-valued, and consider a point in the phase-space \( x^* \) the “outcome” or “prediction” of the model if

\[
f(x^*) = 0 \tag{8}
\]

and some solutions to eq. (7) approach \( x^* \) and \( t \to \infty \). Note that it is possible for there to exist more than one such \( x^* \) for a single model (and parameter set), a condition known as bi-stability[27]. In this case, we consider all such \( x^* \) to be model outcomes.

For details on this process, see A.1.

In Friedman et al.[16], model outcome is defined using direct simulation rather than asymptotic stability.

4.3 Qualitative effect on growth

We use pair growth experiments to determine the qualitative effect of one species on another. That is, we find the difference between the time-average per-organism growth rate of species \( i \) alone and in various pairs or trios.

We label the difference between per-organism growth rate in species \( i \) when grown in some set of species and per-organism growth rate in species \( i \) when grown alone as

\[
\Delta g^\circ_{ij} = \text{mean}\left( \left\{ \frac{x^{\text{with}(o)}(t + \Delta t) - x^{\text{with}(o)}(t)}{x^{\text{with}(o)}(t)} \right\} \right) - \text{mean}\left( \left\{ \frac{x^{\text{alone}}(t + \Delta t) - x^{\text{alone}}(t)}{x^{\text{alone}}(t)} \right\} \right)
\]

where, for example \( o \) is \( j \) for species \( i \) grown a pair with species \( i \), and \( \cdot \) is \( j,k \) for species \( i \) grown in a trio with \( j \) and \( k \). Then, we observe that the additivity of eq. (1) and the assumption that the functions \( h_{ij}, h_{ik} \) do not switch sign imply that, if the microbiota grows according to eq. (1), then

\[
\Delta g^\circ_{ij} > 0 \& \Delta g^\circ_{ik} > 0 \Rightarrow \Delta g^{ik}_{ij} > 0 \tag{9}
\]

and

\[
\Delta g^\circ_{ij} < 0 \& \Delta g^\circ_{ik} < 0 \Rightarrow \Delta g^{ik}_{ij} < 0 \tag{10}
\]

In other words, if two species have the same qualitative effect on the growth of a third, their combination will have that same qualitative effect. We simply compare these quantities in the data to find examples for which this does not hold. We call such an instance a “reversal”. Note that the serial dilution of the growth experiment mean that a reversal of effect is not the result of crowding.
4.4 Generalized Lotka-Volterra parameter fitting

We use non-linear least squares procedures to fit parameters of the gLV model to the time-course experimental data. Additionally, we use a pseudo-genetic algorithm to attempt to find a single set of parameters $\{\alpha_{ij}\}$ which cause the model to recapitulate the growth experiment outcomes.

We fit a set of parameters from eq. (2) to the time-course data. For individual growth parameters $(r_i, K_i$ in eq. (2)), we use individual growth data, and non-linear least squares (implemented in python package scipy.optimize) to fit a logistic curve. We fit the parameters $\alpha_{ij}$ to pair growth experiment data. We again using non-linear least squares (implemented in python package scipy.optimize), computing solution curves for a given parameter set numerically in order to compute residuals. The parameters fitted, along with code to reproduce the fitting, can be found in the supplemental material.

We also ask whether or not the model has the capacity to explain all of the growth experiment outcomes with a single parameter set by searching for such a parameter set. Notice that the interdependence of the trios mean that this is a stricter condition than the existence of a parameter set for each trio which correctly predicts the experimental outcome of just that trio, and also stricter than the existence of a single such parameter set for each trio and involved pairs.

We take advantage of qualitative analysis, detailed in Appendix A, to search for parameter values that explain all of the outcomes observed. To perform this search, we begin with parameter sets that explain 12 independent trios. We then use a pseudo-genetic algorithm, assessing the fitness of the parameter set by the magnitude of the eigenvalues of the various Jacobian matrices whose sign do not match the eigenvalues that would result in a model matching the observed data, to search for a parameter set which allows the model to recapitulate the experiments. In this algorithm, genes can be mutated with a continuous random variable, whereas in a standard genetic algorithm genes are generally taken over a discrete set. See Appendix A and supplementary materials for explanation and code, respectively. The parameters found can be found in the supplemental material.

We next relax the search condition by attempting to find for each pair of species parameters that gave an accurate prediction of all the trios they are involved in without changing the interaction parameters between that pair. This is done in order to identify pairwise relationships that are consistent across different trios. To do this for a pair of species $i, j$, we first fix $\alpha_{ij}$ and $\alpha_{ji}$ as fitted to time-course data. Then, for each other species $k$, we seek $\alpha_{ik}, \alpha_{jk}, \alpha_{ki}, \alpha_{kj}$ so that model asymptotic stability matches the observed outcome of the experiments. We again use a pseudo-genetic algorithm, taking advantage of some partial conditions for equilibrium stability given in the supplementary code. We record for each pair the proportion of trios for which parameters can be found to explain the experimental outcomes. The result is shown in fig. 4.

4.5 The stochastic generalized Lotka-Volterra model

We choose the stochastic process which has the property that as we increase the concentration of organisms modeled we recover the deterministic generalized Lotka-Volterra model. This property means that rather than simply “adding noise” to the model, we have chosen the closest fully stochastic analogue to the model, which is also the standard choice of stochastic model for interactions between agents under the assumption of mass action kinetics. See A.4 for details of the model.

We produce (exact) realizations of the stochastic process using Gillespie’s algorithm (sometimes called the Stochastic Simulation Algorithm), as shown in fig. 8. Additionally, we perform Monte Carlo experiments to estimate the probability (±6%) of each observed outcome according to the stochastic model with parameters fitted to pairs. We use the $\tau$-leaping algorithm from [57] for efficiency.
4.6 Construction of QSMI Model

Starting with parameters given by models of individual growth, we can arrive at every possible pair outcome by assuming there are at most \( \binom{2}{k} \) + 1 external metabolites used for growth, but only one is initially present. We build this model by adding cross-talk molecules to models of growth on a single nutrient. In our model, these cross-talk molecules are produced by one microbe and have some effect on another.

We do this systematically by adjusting pair or trio models separately. This means that we add a cross-talk pathway to a pair model so that it recapitulates the correct growth experiment outcome, and does so in such a way that the added cross-talk pathway will have no effect on any other pair, and likewise with trios. That is, any cross-talk pathway must require all members of the pair or trio it was added for to be present to have some effect. This restriction makes the construction systematic, but is not necessary.

We begin with a model of growth:

\[
\frac{dx_1}{dt} = \kappa_{11}x_1y_1 - d_1x_1 \\
\frac{dy_1}{dt} = f_1 - d_1^*y_1 - \kappa_{11}x_1y_1.
\]

For each pair model, if the initial model formed simply by including both microbes does not match growth outcome, we add a metabolite to the model. Restricted to a pair, the model is then

\[
\frac{dx_1}{dt} = \kappa_{11}x_1y_1 - d_1x_1 + \psi_{12}x_1y_2 \\
\frac{dx_2}{dt} = \kappa_{21}x_2y_1 - d_2x_2 \\
\frac{dy_1}{dt} = f_1 - d_1^*y_1 - \kappa_{11}x_1y_1 - \kappa_{21}x_2y_1 \\
\frac{dy_2}{dt} = \kappa_{21}x_2y_1 - d_2^*y_2 - \kappa_{12}x_1y_2
\]

where \( y_2 \) serves as the cross-talk molecule, in this example produced by \( x_2 \) and having some effect (determined by \( \psi_{12} \)) on \( x_1 \).

To adjust the model to account for trio growth experiments, we add trio-specific cross-talk pathways. Precisely, given a model of organisms \( x_1, x_2, x_3 \) (assuming without loss of generality that \( x_1 \) survives) which may already include pair-wise cross-talk and cross-poisoning, we can introduce a trio specific chain by introducing two new metabolites \( y_1^*, y_2^* \) such that

\[
\frac{dy_1^*}{dt} = \kappa_{11}x_1y_1 - d_1^{*1}y_1^* - \kappa_{21}x_2y_1^* \\
\frac{dy_2^*}{dt} = \kappa_{21}x_2y_1^* - d_2^{*2}y_2^* - \kappa_{12}x_3y_2^*
\]

where \( y_2^* \) has some effect on \( x_3 \). These equations model a situation in which \( y_1^* \) is produced as a metabolic by-product of \( x_1 \) metabolizing \( y_1 \), and likewise \( y_2^* \) is produced as a metabolic by-product of \( x_2 \) metabolizing \( y_2^* \). Then, \( x_1 \) and \( x_2 \) must both be present for \( y_2^* \) to be produced. This chain is specific to this trio as long as \( y_2^* \) only has an effect on \( x_3 \).

We must also for one trio change the model so that it predicts coexistence of the entire trio rather than two extinctions. This requires a signalling molecule and a positive feedback loop of cross-feeding among the two species which were previously predicted to go extinct. This model therefore needs three additional molecules.

Finally, there is one trio for which we need the model to predict one lone survivor instead of a different lone survivor. This requires a signal that is not consumed or diluted (or was diluted at a very small rate). To do this, we use two signalling molecules that are never degraded or consumed, and cause one species to poison the other two species with a third molecule.

The complete model is detailed in the supplemental material, and a schematic of the interactions in the model is shown in fig. 6.

4.7 Estimation of QSMI model complexity

To estimate QSMI model complexity, we reduce our set of allowed models to only those with one initial available metabolite and direct cross-talk between pairs as well as cross-talk chains as detailed in eqs. (17) and (18). This allows us to automatically generate a model which explains a large proportion (98% in the case of the data from Friedman et al.\cite{16}) of a set of pair and trio growth experimental outcomes. We attempt to maximize this coverage and minimize the number of molecules needed in order to estimate the complexity.
of QSMI models. In our construction, the number of additional pathways added is ultimately a function of the microbes relative ability to metabolize the initial available metabolite. This observation allows us to optimize over the set of possible orders of these metabolic parameters.

We generate random trio and growth experiment outcomes by permuting the real experimental outcomes. In this way, we preserve the number of each possible qualitative outcome (i.e. coexistence, extinction, and double extinction). We then generate the best possible model for each random outcome set.

5 Discussion

We would like to build a clinically useful model of the dynamics of the human microbiome. For this, we seek a modeling framework that infers community dynamics from fundamental interactions, so that data and discoveries from across studies can be incorporated into an individualized model using these interactions as building blocks. We therefore need a model that can be built without reparameterization and can capture emergent properties of microbial community composition dynamics.

As a representative example of species-species modeling, we inspect the generalized Lotka-Volterra model. Using this model, we see disagreement between trio growth experiments and model prediction based on pairwise fitted parameters, and we are unable to find a set of parameters which allow the model to recapitulate the qualitative outcomes of the interdependent growth experiments. This suggests that the generalized Lotka-Volterra model has a high sensitivity to fitted parameters, even for qualitative results, and that the space of parameters which fit the entire set of qualitative growth experiments is small or does not exist.

In species-metabolite modeling, dynamics are modeled by the interactions of individual microbes with a shared metabolite pool. We find that this framework has the additional complexity necessary for capturing emergent behavior through cross-feeding and cross-poisoning. We show that this framework does not adhere to the additive interaction assumption, and that a model can be found to fit interdependent growth data. We then use the mechanisms of cross-feeding and cross-poisoning to fit various competitive configurations to complex outcomes.

It is worth noting that SMI models are inherently more complex than SSI models, as proven by Momeni et al.[11]. Our result, which relates SSI and SMI modeling to experimental growth data, compliments the conclusion of Momeni et al.[11] by showing that the dynamical complexity is indeed necessary for accurate modeling. The appeal of SMI models, however, goes beyond mere complexity. SMI models may provide better fundamental building blocks for inferring community dynamics because they better reflect the real biological building blocks of microbial community interactions. This allows us to build models for larger communities by simply combining models for smaller communities. In contrast, one might add higher order terms to an SSI model to account for interactions between more than two species. However, this approach requires that a model built for a large community has little relationship with models of sub-communities. Such an extension of SSI models therefore does not achieve our goal of individualized modeling.

Our work focuses on establishing the capacity of different modeling frameworks for recapitulating the emergent behavior of relatively simple microbial communities. While this work provides an important starting point, it is nonetheless limited in applicability. For instance, it is fair to say that the procedure to build the QSMI model for the growth experiments in Friedman et al.[16] is not in the spirit of building a model in an individualized manner. In practice, cross-feeding and other interactions of the microbe with the metabolite pool might be discovered from, for example, metabolic modeling, and SMI models may be built from genome-scale metabolic models of individual microbial species[31, 58, 59]. In this way, high-throughput sequencing technology can be leveraged to better understand microbiome composition. Developing methods to build clinically useful species-metabolite models with available data remains an open and interesting area of research.

In addition, we investigate the complexity of the QSMI model by minimizing the number of metabolites needed to match most of a set of growth experiment outcomes. It would also be interesting to build the QSMI model with further “reality” criteria in mind, such as using the minimum number of cross-talk pathways. In future work, we plan to establish a systematic method to build a QSMI model which matches some set of outcomes exactly and satisfies various reality conditions. It is also of interest to establish easily checkable conditions on a set of growth experiment outcomes which decide whether or not there exists a QSMI which can recapitulate the experiments. While these questions are very interesting, they are fundamentally
mathematical considerations and outside of the scope of this paper, which attempts to determine model usefulness in the context of real data.

Species-metabolite interaction models provide an intermediate level of complexity between fully detailed genome-scale models and fully simplified models such as species-species interaction models. This level of complexity holds promise for individualized predictions in medicine.

This work was supported by funding from the Andersen Family Foundation, National Cancer Institute grant R01 CA179243, and the Center for Individualized Medicine, Mayo Clinic.

Code for parameter estimation and searching, parameter values, and a complete description of the QSMI model, as well as code for the stochastic experiments, is available at https://github.com/jdbrunner/model_comparisons.

JB carried out computational, statistical, and mathematical analysis, and drafted the manuscript. NC directed the goals of the analysis, and critically revised the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

References

[1] Braundmeier AG, Lenz KM, Inman KS, Chia N, Jeraldo P, Walther-António MRS, Berg Miller ME, Yang F, Creedon DJ, Nelson H, White BA. 2015 Individualized medicine and the microbiome in reproductive tract. *Frontiers in Physiology* 6, 97.

[2] Calcinoatto A, Brevi A, Chesi M, Ferrarese R, Perez LG, Grioni M, Kumar S, Garbitt VM, Sharik ME, Henderson KJ et al.. 2018 Microbiota-driven interleukin-17-producing cells and eosinophils synergize to accelerate multiple myeloma progression. *Nature communications* 9, 4832.

[3] Walsh DM, Mert I, Chen J, Hou X, Weroha SJ, Chia N, Nelson H, Mariani A, Walther-Antonio MR. 2019 The Role of Microbiota in Human Reproductive Tract Cancers. In *AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY* vol. 168 pp. 260–261. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA.

[4] Hale VL, Jeraldo P, Chen J, Mundy M, Yao J, Priya S, Keeney G, Lyke K, Ridlon J, White BA, French AJ, Thibodeau SN, Diener C, Resendis-Antonio O, Gransee J, Dutta T, Petterson XM, Sung J, Blekhman R, Boardman L, Larson D, Nelson H, Chia N. 2018 Distinct microbes, metabolites, and ecologies define the microbiome in deficient and proficient mismatch repair colorectal cancers. *Genome Medicine* 10, 78.

[5] Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, O’riordain M, Shanahan F, O’toole PW. 2017 Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut* 66, 633–643.

[6] Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, Naidu N, Choudhury B, Weimer BC, Monack DM, Sonnenburg JL. 2013 Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* 502, 96 EP –.

[7] Round JL, Mazmanian SK. 2009 The gut microbiota shapes intestinal immune responses during health and disease. *Nature Reviews Immunology* 9, 313 EP –.

[8] Chen J, Chia N, Kalari KR, Yao JZ, Novotna M, Paz Soldan MM, Luckey DH, Marietta EV, Jeraldo PR, Chen X, Weinschenker BG, Rodriguez M, Kantarci OH, Nelson H, Murray JA, Mangalam AK. 2016 Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. *Scientific Reports* 6, 28484 EP –.

[9] Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. 1990 The n-of-1 randomized controlled trial: clinical usefulness: our three-year experience. *Annals of internal medicine* 112, 293–299.
[10] Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. 2011 The n-of-1 clinical trial: the ultimate strategy for individualizing medicine?. *Personalized medicine* 8, 161–173.

[11] Momeni B, Xie L, Shou W. 2017 Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions. *Elife* 6, e25051.

[12] Wang T, Goyal A, Dubinkina V, Maslov S. 2019 Evidence for a multi-level trophic organization of the human gut microbiome. *bioRxiv* p. 603365.

[13] Erez A, Lopez JG, Weiner B, Meir Y, Wingreen NS. 2019 Nutrient levels and trade-offs control diversity in a model seasonal ecosystem. *arXiv preprint arXiv:1902.09039*.

[14] Goyal A, Maslov S. 2018 Diversity, Stability, and Reproducibility in Stochastically Assembled Microbial Ecosystems. *Phys. Rev. Lett.* 120, 158102.

[15] Goyal A, Dubinkina V, Maslov S. 2017 Microbial community structure predicted by the stable marriage problem. *Preprint at http://arxiv. org/abs/1712.06042*.

[16] Friedman J, Higgins LM, Gore J. 2017 Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution* 1, 0109 EP –.

[17] Mounier J, Monnet C, Vallaeys T, Ardit R, Sarthou AS, Hélias A, Irlinger F. 2008 Microbial interactions within a cheese microbial community. *Appl. Environ. Microbiol.* 74, 172–181.

[18] Fisher CK, Mehta P. 2014 Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PloS one* 9, e102451.

[19] Stein RR, Bucci V, Toussaint NC, Buffie CG, Rättsch G, Pamer EG, Sander C, Xavier JB. 2013 Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS computational biology* 9, e1003388.

[20] Kuntal BK, Gadgil C, Mande SS. 2019 Web-gLV: A web based platform for Lotka-Volterra based modeling and simulation of microbial populations. *Frontiers in microbiology* 10.

[21] Angulo MT, Moog CH, Liu YY. 2019 A theoretical framework for controlling complex microbial communities. *Nature communications* 10.

[22] Faust K, Raes J. 2012 Microbial interactions: from networks to models. *Nature Reviews Microbiology* 10, 538 EP –.

[23] Müller H, Mancuso F. 2008 Identification and Analysis of Co-Occurrence Networks with NetCutter. *PLOS ONE* 3, 1–16.

[24] Feinberg M. 1979 Lectures on Chemical Reaction Networks. .

[25] Yu PY, Craciun G. 2018 Mathematical analysis of chemical reaction systems. *Israel Journal of Chemistry* 58, 733–741.

[26] Billick I, Case TJ. 1994 Higher Order Interactions in Ecological Communities: What Are They and How Can They be Detected?. *Ecology* 75, 1529–1543.

[27] Edelstein-Keshet L. 2005 *Mathematical Models in Biology* vol. 46*Classics in Applied Mathematics*. SIAM.

[28] Niehaus L, Boland I, Liu M, Chen K, Fu D, Henekel C, Chaung K, Espinoza Miranda S, Dyckman S, Crum M, Dedrick S, Shou W, Momeni B. 2018 Microbial coexistence through chemical-mediated interactions. *bioRxiv*.

[29] Posfai A, Taillefumier T, Wingreen NS. 2017 Metabolic Trade-Offs Promote Diversity in a Model Ecosystem. *Phys. Rev. Lett.* 118, 028103.
[30] Sung J, Kim S, Cabatbat JJT, Jang S, Jin YS, Jung GY, Chia N, Kim PJ. 2017 Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nature communications* **8**, 15393; 15393–15393.

[31] Chan SHJ, Simons MN, Maranas CD. 2017 SteadyCom: Predicting microbial abundances while ensuring community stability. *PLOS Computational Biology* **13**, 1–25.

[32] Hart SFM, Skelding D, Waite AJ, Burton J, Xie L, Shou W. 2018 Microscopy quantification of microbial birth and death dynamics. *BioRxiv* p. 324269.

[33] Diener C, Resendis-Antonio O. 2018 Micom: metagenome-scale modeling to infer metabolic interactions in the microbiota. *bioRxiv*.

[34] Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. 2010 High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology* **28**, 977 EP –.

[35] Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. 2013 COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology* **7**, 74.

[36] Röttjers L, Faust K. 2018 From hairballs to hypotheses — biological insights from microbial networks. *FEMS Microbiology Reviews* **42**, 761–780.

[37] Mougi A, Kondoh M. 2012 Diversity of interaction types and ecological community stability. *Science* **337**, 349–351.

[38] Thébault E, Fontaine C. 2010 Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329**, 853–856.

[39] Allesina S, Tang S. 2012 Stability criteria for complex ecosystems. *Nature* **483**, 205.

[40] Dohlman AB, Shen X. 2019 Mapping the microbial interactome: Statistical and experimental approaches for microbiome network inference. *Experimental Biology and Medicine* p. 1535370219836771.

[41] Shaw GTW, Pao YY, Wang D. 2016 MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles. *Bmc Bioinformatics* **17**, 488.

[42] Chen WY, Ng TH, Wu JH, Chen JW, Wang HC. 2017 Microbiome dynamics in a shrimp grow-out pond with possible outbreak of acute hepatopancreatic necrosis disease. *Scientific reports* **7**, 9395.

[43] Shaw GTW, Liu AC, Weng CY, Chou CY, Wang D. 2017 Inferring microbial interactions in thermophilic and mesophilic anaerobic digestion of hog waste. *PloS one* **12**, e0181395.

[44] Džunková M, Martinez-Martinez D, Gardlík R, Behuliak M, Janšáková K, Jiménez N, Vázquez-Castellanos JF, Martí JM, DAuria G, Bandara H et al.. 2018 Oxidative stress in the oral cavity is driven by individual-specific bacterial communities. *npj Biofilms and Microbiomes* **4**, 29.

[45] Feinberg M. 1979 Lectures on Chemical Reaction Networks. .

[46] Gause GF. 1934 *The struggle for existence*. Baltimore,The Williams & Wilkins company,.

[47] Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM et al.. 2012 Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences* **109**, E1743–E1752.

[48] Pérez-Cobas AE, Gosalbes MJ, Friedrichs A, Knecht H, Artacho A, Eismann K, Otto W, Rojo D, Bargiela R, von Bergen M et al.. 2013 Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut* **62**, 1591–1601.
A The Positive Steady State of the generalized Lotka-Volterra model

A.1 Asymptotic stability analysis

We can re-scale eq. (2) with two species to

\[
\frac{dx_1}{dt} = r_1 x_1 (1 - x_1 + \alpha_{12} x_2) \\
\frac{dx_2}{dt} = r_2 x_2 (1 - x_2 + \alpha_{21} x_1)
\]

in order to simplify notation, and analyze asymptotic behavior of this model by performing straightforward stability analysis on the equilibrium\[27\]. We see that eq. (19) has equilibrium at \((0, 0)\), \((1, 0)\), \((0, 1)\) and \(\mathbf{x}^* = \left( \frac{1 + \alpha_{12}}{1 - \alpha_{12} \alpha_{21}}, \frac{1 + \alpha_{21}}{1 - \alpha_{12} \alpha_{21}} \right)\).

Furthermore, linearization about each of those points reveals that \((0, 0)\) is never stable, \((1, 0)\) is stable if \(\alpha_{21} < -1\), \((0, 1)\) is stable if \(\alpha_{12} < -1\). Lastly, \(\mathbf{x}^* \in \mathbb{R}_+^2\) if and only if \(\{\alpha_{12} < -1, \alpha_{21} < -1\}\) or \(\{-1 < \alpha_{12}, -1 < \alpha_{21}, \alpha_{12} \alpha_{21} < 1\}\) and if \(\{\alpha_{12} < -1, \alpha_{21} < -1\}\), then \(\mathbf{x}^*\) is unstable, and is in fact a saddle point. If \(\{-1 < \alpha_{12}, -1 < \alpha_{21}, \alpha_{12} \alpha_{21} < 1\}\), then \(\mathbf{x}^*\) is stable. All of this can be done through symbolic analysis of the Jacobian matrix evaluated at \(\mathbf{x}^*\).

We can now characterize the outcomes observed in the paper using the parameters \(\alpha_{12}\) and \(\alpha_{21}\):
(a) Coexistence: this is stability of the positive state, and so requires \{-1 < \alpha_{12}, -1 < \alpha_{21}, \alpha_{12}\alpha_{21} < 1\}.

(b) Invasion of one species regardless of initial condition: this is stability of one boundary state and instability of the other. This requires \(\alpha_{12} < -1, \alpha_{21} > -1\) or the opposite. If \(\alpha_{12} < -1\) then 2 invades 1.

(c) Bi-stability: This is stability of both boundary states, and requires \(\{\alpha_{12} < -1, \alpha_{21} < -1\}\). Interestingly, case (c) is not observed in the data of [16].

The three species model is

\[
\begin{align*}
\frac{dx_1}{dt} &= r_1x_1(1 - x_1 + \alpha_{12}x_2 + \alpha_{13}x_3) \\
\frac{dx_2}{dt} &= r_2x_2(1 - x_2 + \alpha_{21}x_1 + \alpha_{23}x_3) \\
\frac{dx_3}{dt} &= r_3x_3(1 - x_3 + \alpha_{31}x_1 + \alpha_{32}x_2)
\end{align*}
\]

and here again we can compute model equilibrium states and stability. There are 8 equilibrium points, corresponding to each qualitative possibility of survival & extinction. Again, the \((0,0,0)\) equilibrium is never stable. There exist simple conditions on the parameters for local stability of all equilibrium points except for the state which represents coexistence of all three microbes. Stability of this last state can, however, be easily evaluated for any given parameters.

We can compute the Jacobian determinant to see that the stability conditions for the double extinction equilibrium points are

- \((1,0,0)\) : \(\alpha_{21} < -1, \alpha_{31} < -1\)
- \((0,1,0)\) : \(\alpha_{12} < -1, \alpha_{32} < -1\)
- \((0,0,1)\) : \(\alpha_{13} < -1, \alpha_{23} < -1\)

Taking advantage of symmetry, we investigate only one of the three single extinction equilibrium, which have the form \((\frac{1+\alpha_{12}}{1-\alpha_{12}\alpha_{21}}, \frac{1+\alpha_{21}}{1-\alpha_{12}\alpha_{21}}, 0)\). The first two eigenvalues of the Jacobian matrix at these points will follow the two dimensional case, so we have the necessary conditions for stability \(-1 < \alpha_{12}, -1 < \alpha_{21}, \alpha_{12}\alpha_{21} < 1\). This is simply because after extinction of species \(k\), the model is identical to the pair model. While unsurprising, this fact does imply that not all hypothetical combinations of existence and extinction outcomes for pair and trio experiments can be simultaneously explained by the parameters of the generalized Lotka-Volterra model. However, there were no instances in the trio experiments being considered in which such a “smoking gun” scenario was observed.

The third eigenvalue is the value of \(r_3(1 - 2x_3 + \alpha_{31}x_1 + \alpha_{32}x_2)\) evaluated at this point, which is

\[
\lambda_3 = r_3 \left( 1 + \alpha_{31} \frac{1 + \alpha_{12}}{1 - \alpha_{12}\alpha_{21}} + \alpha_{32} \frac{1 + \alpha_{21}}{1 - \alpha_{12}\alpha_{21}} \right)
\]

Clearly if \(\alpha_{31}\) and \(\alpha_{32}\) are both positive, this state is unstable. The condition for linear stability is

\[
\alpha_{31} \frac{1 + \alpha_{12}}{1 - \alpha_{12}\alpha_{21}} + \alpha_{32} \frac{1 + \alpha_{21}}{1 - \alpha_{12}\alpha_{21}} < -1.
\]

**A.2 Lack of limit cycles of the two-species gLV model**

We can rule out closed orbits in the two species gLV model using Dulac’s criterion. Letting

\[
g(x_1, x_2) = \frac{1}{x_1x_2}
\]

we compute

\[
\nabla \cdot \left( \frac{dx_1}{dt}, \frac{dx_2}{dt} \right)g(x_1, x_2) = \frac{r_1}{x_2} - \frac{r_2}{x_1} < 0
\]
for all $x_1, x_2 > 0$. This implies that there are no solution to eq. (19) is a closed orbit in the positive quadrant.

Note that the standard predator-prey “Lotka-Volterra” model does allow closed orbits. This is because that model does not include the quadratic terms $-r_1 x_1^2$ and $-r_2 x_2^2$ that appear in eq. (19), and because that model does not enforce the assumption that $r_i > 0$. This can be interpreted as an assumption of infinite carrying capacity of the prey species and exponential decay of the predator.

A.3 Pseudo-Genetic Algorithm

We search for a parameter set to fit qualitative growth behavior by performing a pseudo-genetic algorithm which attempts to minimize

$$M = \sum_{\text{trios}} \left( \sum_{\lambda_i} \lambda^* + p_i \right)$$

(27)

where $\Lambda_i$ is the set of eigenvalues which corresponded to the equilibrium point which matches the experimental outcome of trio $i$, and $p_i = 0$ if the three pairs of trio $i$ match experimental outcome, and $p_i = 1000$ otherwise. The chromosomes of the genetic search are taken to be the parameter sets, represented as a matrix whose $i,j$ entry contained $\alpha_{ij}$. We use the rows of this matrix as genes, and so the mating procedure is to choose for each row of the child the row of one or the other parent with even probability.

We describe this as a “pseudo-genetic” algorithm because we are searching over a continuous parameter space. In order to account for this, random mutation of parameters is done by perturbation with a continuous random variable. First, to determine if mutation occurred, we draw a uniform random variable in $(0, 1)$ and mutate if this variable is less than a thresh-hold of 0.2. If mutation occurs, a random matrix whose entries are generated uniformly in $[-0.05, 0.05]$ is added to the matrix representing the parameter set.

A.4 The stochastic generalized Lotka-Volterra model

The model is as follows:

$$X_i(t) = X_i(0) + Y_{i1}^1 \left( r_i \int_0^t X_i(s) ds \right)$$

$$- Y_{i2}^2 \left( \hat{r}_i \int_0^t X_i(s)(X_i(s) - 1) ds \right)$$

$$+ \sum_{i \neq j} Y_{ij} \left( \hat{\alpha}_{ij} \int_0^t X_i(s) X_j(s) ds \right).$$

(28)

Here, $Y(p(t))$ are non-homogeneous Poisson (jump) processes with time-varying propensity $p(t) = \int_0^t f(s) ds$. The new parameters $\hat{r}_i$ and $\hat{\alpha}_{ij}$ depend on the “volume” of the experiment, i.e. the population size scale. Precisely, with a volume $N$ we take $\alpha_{ij}$ as fitted to pair growth experiments and let

$$\hat{r}_i = \frac{r_i}{N}, \quad \hat{\alpha}_{ij} = \frac{\alpha_{ij}}{N}$$

Then, as $N \to \infty$, realizations of the stochastic model $\frac{X}{N}$ approach trajectories of the deterministic model [32].
B Stability of equilibrium of QSMI model.

Consider the model for \( n \) microbes

\[
\frac{d}{dt}x_i = \kappa_i x_i y - d_i x_i \quad i = 1, \ldots, n \tag{29}
\]

\[
\frac{d}{dt}y = f_y - d_y y - \sum_{i=1}^{n} \kappa_i x_i y \tag{30}
\]

This has equilibrium at \( x_i = 0, y = \frac{f_y}{d_y} \), and at \( y = \frac{d_i}{\kappa_i} \) for each \( i \), with \( x_j > 0 \) if and only if \( \frac{d_j}{\kappa_j} = \frac{d_i}{\kappa_i} \). The general form of the characteristic equation of the Jacobian matrix about any steady state for this system is

\[
\det(J - \lambda I) = (-d_y - \sum_{i=1}^{n} \kappa_i x_i - \lambda) \prod_{i=1}^{n} (\kappa_i y - d_i - \lambda) + \sum_{i=1}^{n} \kappa_i^2 y x_i \prod_{j \neq i} (\kappa_j y - d_j - \lambda) \tag{31}
\]

Solving at the extinction steady state \( x_i = 0, y = \frac{f_y}{d_y} \), we have eigenvalues \( \lambda_{n+1} = -d_y, \lambda_j = \frac{\kappa_j f_x}{d_y} - d_j \). Therefore, this state is linearly stable if and only if

\[
d_i d_y > k_i f_y \quad \forall i \tag{32}
\]

Next, for each \( i \) let \( Q_i = \left\{ j \mid \frac{d_j}{\kappa_j} = \frac{d_i}{\kappa_i} \right\} \). Then for each \( i \) we have the set of equilibrium defined by

\[
\sum_{j \in Q_i} \kappa_j x_j = \frac{\kappa_i}{d_i} \left( f_y - d_y \frac{d_i}{\kappa_i} \right) \tag{33}
\]

and \( x_i = 0 \) if \( l \notin Q_i \). The characteristic equation becomes

\[
\det(J - \lambda I) = -\left[ \lambda^2 + \lambda d_y + \frac{\kappa_i}{d_i} \left( f_y - d_y \frac{d_i}{\kappa_i} \right) + \sum_{j \in Q_i} \kappa_j^2 x_j \right] \lambda^{\left| Q_i \right|-1} \prod_{l \notin Q_i} \left( \frac{\kappa_l}{\kappa_i} \frac{d_l}{d_i} - d_l - \lambda \right) \tag{34}
\]

First, we see that if \( \frac{d_i}{\kappa_i} \neq \min_{l=1, \ldots, n} \left\{ \frac{d_l}{\kappa_l} \right\} \), then this is unstable. If we do have the minimum \( \frac{d_i}{\kappa_i} \), then the remaining nonzero eigenvalues are

\[
\lambda_{\pm} = \frac{1}{2} \left[ -B \pm \left( B^2 + 4A \right)^{1/2} \right] \tag{35}
\]

where \( B = d_y + \frac{\kappa_i}{d_i} \left( f_y - d_y \frac{d_i}{\kappa_i} \right) > 0 \) and \( A = \sum_{j \in Q_i} \kappa_j^2 x_j > 0 \). These both then have negative real part, implying that the hyperplane of solutions is attracting (note that if \( |Q_i| = 1 \), this implies a linearly stable equilibrium point).

Next, we consider the two species cross-feeding or cross-poisoning model:

\[
\frac{d}{dt}x_1 = \kappa_{11} x_1 y_1 - d_1 x_1 + \psi_{12} x_1 y_2 \tag{36}
\]

\[
\frac{d}{dt}x_2 = \kappa_{21} x_2 y_1 - d_2 x_2 \tag{37}
\]

\[
\frac{d}{dt}y_1 = f_1 - d_1^* y_1 - \kappa_{11} x_1 y_1 - \kappa_{21} x_2 y_1 \tag{38}
\]

\[
\frac{d}{dt}y_2 = \kappa_{21} x_1 y_1 - d_2^* y_2 - \kappa_{12} x_1 y_2 \tag{39}
\]

Here, conditions for stability of the double extinction state are the same as above. Suppose \( d_2 \kappa_{11} > d_1 \kappa_{21} \), so that if \( \psi_{12} = 0 \), this model behaves as the single metabolite model, and the state with \( x_2 = 0, x_1 > 0 \) is
stable. We are interested in causing the opposite extinction. That steady state is

\[(x_1, x_2, y_1, y_2) = \left(0, \frac{1}{d_2}, \frac{f_1 - d_1^* d_2}{\kappa_{21}}, \frac{d_2}{\kappa_{21}}, \frac{1}{d_2}, \frac{f_1 - d_1^* d_2}{\kappa_{21}} \right)\]

(40)

and the eigenvalues of the Jacobian matrix at this state can be computed symbolically, and the relevant eigenvalue is

\[\lambda = \kappa_{11} \frac{d_2}{\kappa_{21}} - d_1 + \psi_{12} \left(f_1 - d_1^* \frac{d_2}{\kappa_{21}}\right)\]

(41)

giving a condition for stability on \(\psi_{12}\) that can be achieved.

For coexistence, we will assume that the initial model with \(\psi_{12} = 0\) has survival of \(x_2\), so \(d_2 \kappa_{11} < d_1 \kappa_{21}\). Then we simply repeat the argument above to destabilize the equilibrium point, causing \(\lambda > 0\). Then, all three of the double extinction and both single extinction equilibrium are unstable. We can conclude coexistence.