Intrinsic instability of the dysbiotic microbiome revealed through dynamical systems inference at scale

Travis E. Gibson[1,3,6], Younhun Kim[5], Sawal Acharya[1], David E. Kaplan[1], Nicholas DiBenedetto[2], Richard Lavin[2], Bonnie Berger[5,6], Jessica R. Allegretti[3,7], Lynn Bry[2,3], and Georg K. Gerber[1,2,3,4,#]

1 Division of Computational Pathology, Brigham and Women's Hospital, Boston MA, USA
2 Massachusetts Host Microbiome Center, Brigham and Women's Hospital, Boston MA, USA
3 Harvard Medical School, Boston MA, USA
4 MIT-Harvard Health Sciences and Technology, Cambridge MA, USA
5 Mathematics Department, MIT, Cambridge MA, USA
6 Computer Science and Artificial Intelligence Lab, MIT, Cambridge, Boston MA, USA
7 Division of Gastroenterology, Brigham and Women's Hospital, Boston MA, USA
Correspondence: ggerber@bwh.harvard.edu

Abstract: Despite the importance of microbial dysbiosis in human disease, the phenomenon remains poorly understood. We provide the first comprehensive and predictive model of dysbiosis at ecosystem-scale, leveraging our new machine learning method for efficiently inferring compact and interpretable dynamical systems models. Coupling this approach with the most densely temporally sampled interventional study of the microbiome to date, using microbiota from healthy and dysbiotic human donors that we transplanted into mice subjected to antibiotic and dietary interventions, we demonstrate superior predictive performance of our method over state-of-the-art techniques. Moreover, we demonstrate that our approach uncovers intrinsic dynamical properties of dysbiosis driven by destabilizing competitive cycles, in contrast to stabilizing interaction chains in the healthy microbiome, which have implications for restoration of the microbiome to treat disease.

Introduction

Dysbiosis of the gut microbiome, or disruption of the normal composition of the microbes present, has been associated with a variety of human diseases, including infectious[1], neurological/psychiatric[2,3], autoimmune[4], metabolic[5] and malignancies[6]. Cross-sectional studies have typically been employed to find such associations, and have produced large compendia of compositions of microbial taxa in dysbiotic stool samples. However, a major gap remains between cataloging static snapshots of dysbiotic microbiomes and understanding what properties of the complex gut microbial ecosystem drive dysbiosis. Microbes are inherently dynamic[7], changing over time due to both internal interactions as well as responses to external perturbations. The dynamics of a microbiome reveals important information about how microbes interact and how the ecosystem as a whole *behaves*; for instance, unstable responses to perturbations can be indicative of an inability to maintain homeostatic function[1]. Mathematical models of dynamical systems have a long history in ecology and biomedicine, and have led to many insights, including for microbial ecosystems[8]. Dynamical systems models are particularly powerful because, once inferred from data, they can be directly interrogated using mathematical tools or computational simulations to study aspects including: stability and other ecological properties[9-12], topological properties of the interaction network, including motifs[13-15], and *in silico* forecasts of the system, such as "knock-outs" of taxa or responses to perturbations not yet experimentally studied. However, to successfully apply dynamical systems analyses to the gut microbiome two elements are needed: (1) computational models and inference algorithms capable of handling the scale and complexity of microbiome data, and (2) sufficiently rich data, in order to provide the information necessary to infer parameters of the computational models.

Scale and complexity of microbiomes, as well as limitations of measurement modalities, present modeling and analytical challenges. The gut microbiome contains hundreds of ecologically diverse yet interacting microbes. Indeed, there is increasing recognition that this interaction structure is critically important, driving factors such as whether so-called pathobiont bacteria will cause disease or remain harmless in the host[16]. A well-established modeling framework, which we and others have employed, uses the generalized Lotka-Volterra (gLV) equations[17-20] to model pair-wise interactions among microbial taxa. Although gLV models have been shown empirically to predict microbial dynamics with good accuracy for small ecosystems[19], these models present significant challenges for scalability and interpretability, because the number of modeled interactions increases quadratically with the number of taxa in the system (e.g., for a system of 300 taxa, 89,700 interactions must be modeled). All these model parameters must ultimately be inferred from microbiome data, which itself has fundamental limitations. The frequency and regularity of sampling is dependent on gut transit time and practical logistics, particularly for human studies. Further, microbiome data relies on sequencing and other high-throughput methods, which have complicated noise characteristics. Measurement noise presents particular challenges when we seek to characterize low abundance components of the microbiome, which can serve critical ecological roles [21,22], but are orders of magnitude lower than the predominant taxa in the gut.

Sufficiently rich experimental data is critical for inferring dynamical systems models of the gut microbiome. Perturbations are essential to analyze how components of dynamical systems *interact* and to assess the stability of systems; data at equilibrium in a single or small number of conditions cannot be used to infer these properties of the system. Culture experiments provide one means to introduce controlled perturbations. Traditional approaches studying single microbial isolates in solid or liquid culture systems have yielded deep insights into bacterial genetics and functions, but are not well-suited for studying complex ecosystems. More recent higher-throughput *in vitro* approaches include multi-stage fermenters[23], microfluidic devices[24] and microbial-host cell co-culture systems[25]. Even with these advances, culturing native

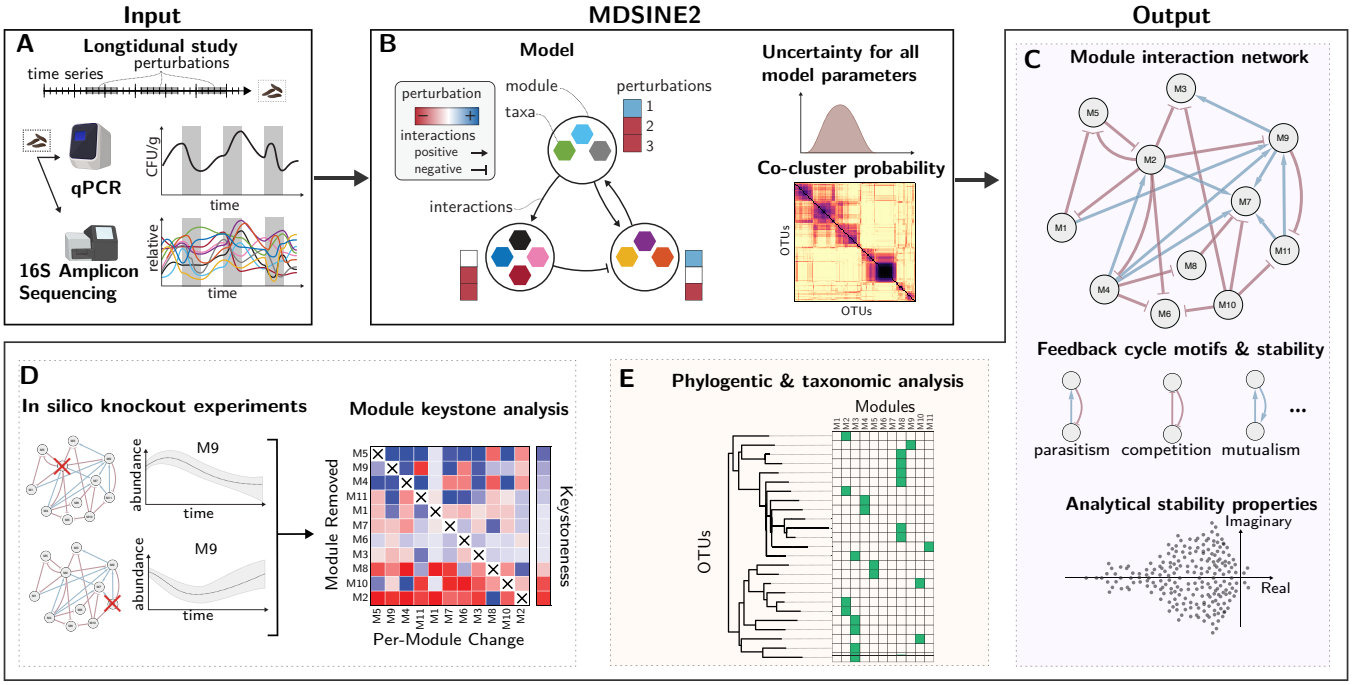

Figure 1. Schematic of the MDSINE2 computational method for inferring interpretable dynamical systems models of microbiomes at scale. (A) Input data to the method are measurements of total bacterial concentration obtained via qPCR, and measurements of taxa abundances obtained via 16S rRNA amplicon sequencing. These measurements are obtained from studies in which the microbiome undergoes perturbations, designed to provide sufficiently rich information to infer dynamical systems models. **(B)** MDSINE2 infers interpretable dynamical systems models by automatically learning interaction modules, or groups of taxa that share the same interactions with other modules and perturbations. The method is fully Bayesian and propagates error throughout the model, providing estimates of uncertainty for all variables including module memberships (co-cluster probabilities). The software provides a variety of tools for analysis and visualization of the inferred dynamical system, including: **(C)** microbial interaction network structure and stability, **(D)** keystoneness (quantitative impact on the ecosystem when modules are removed), and **(E)** module-level analyses of taxonomic composition and phylogeny.

microbiomes *in vitro* remains very challenging, especially for longer durations of time, due to the extremely varied chemical, physical and nutritional requirements of the diverse micro-organisms present. Gnotobiotic mice, in which bacteria-free mice are inoculated with micro-organisms, are a compelling experimental system that has been shown to sustain the vast majority of bacterial species found in native human microbiomes[26] and recapitulate important aspects of human physiology and pathology[19,27,28]. From the standpoint of dynamical systems inference, gnotobiotic models allow for targeted *strong* perturbations, such as dietary changes or antibiotics. Gnotobiotic models also allow the experimenter to limit unwanted external perturbations or variability, including the genetic background of the host, and thus isolate properties that are *intrinsic* to the microbiome, rather than due to uncontrolled interactions with the environment or host genetics.

To analyze microbial dysbiosis at ecosystem-scale, and address the challenges described above, we developed a new computational method, MDSINE2, and generated (to our knowledge) the densest interventional time-series study to date. The remainder of this paper is structured as follows. First, we demonstrated MDSINE2's underlying computational innovations that enable scalable, interpretable, and accurate inference from microbiome time-series data. Second, we characterized microbiome dynamics in our experiments, of colonization of two cohorts of gnotobiotic mice with human microbiota from a healthy or dysbiotic donor (an average of 77 fecal samples/mouse over a 65-day period). Third, we compared performance of MDSINE2 against other state-of-the-art computational methods in forecasting dynamics on our dataset, and demonstrated superior performance of our method. Fourth and finally, we demonstrated that our combined computational and -experimental approach provides the first comprehensive and predictive model of dysbiosis at ecosystem-scale, revealing intrinsic instability of the dysbiotic microbiome driven by chains of ecological interactions and keystone taxa automatically identified by our method.

Results

MDSINE2 is a computational framework for inferring dynamical systems models of microbiomes at scale

To infer accurate and interpretable dynamical systems models from microbiome time-series data, we developed MDSINE2 (Figure 1), a fully Bayesian machine learning model and associated software package that performs inference and provides tools for analysis and visualization of results. The inputs to the MDSINE2 software are time-series measurements of relative bacterial abundances and total bacterial concentrations, derived from 16S rRNA amplicon sequencing and qPCR with universal 16S rRNA primers, respectively. The software then provides a variety of tools for interrogating the inferred dynamical systems model, including forecasting trajectories of taxa (including under perturbations or with removal of taxa from the system), analyzing topological properties of the interaction network, quantitating the keystoneness of individual taxa or modules, and formally assessing the stability of the microbial ecosystem.

Our computational model introduces several innovations. First, MDSINE2 extends the generalized Lotka-Volterra (gLV) model to include automatically learned *interaction modules*, which we define as groups of taxa that share common interaction structure (i.e., are

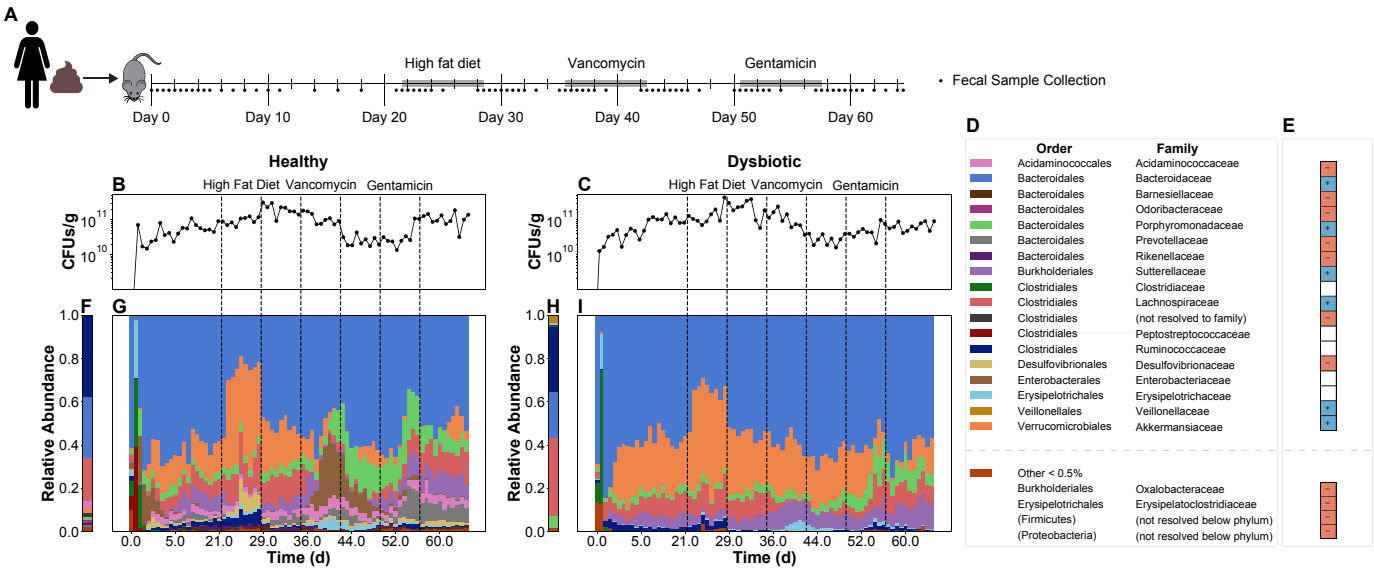

Figure 2: High-temporal resolution gnotobiotic mice colonization and perturbation studies using healthy and dysbiotic human donor microbiomes show persistent but differential colonization. (A) Experimental design for studies (n = 4 mice in healthy donor cohort; n = 5 mice in dysbiotic donor cohort) with 77 serial fecal samples/mouse. **(B)** Average total bacterial concentrations in serial fecal samples from mouse cohort receiving healthy donor microbiome. **(F)** Relative abundance of microbes in starting inoculum from the healthy donor. **(G)** Relative abundances of microbes in serial fecal samples from mouse cohort receiving healthy donor microbiome, averaged over the biological replicates. **(C)**, **(H)**, **(I)** corresponding data from mouse cohort receiving dysbiotic donor microbiome. **(D)** Legend for figures (F), (G), (H), and (I). **(E)** Abundances of seventeen taxa at the family level differed significantly at steady state between gnotobiotic mice that received healthy versus dysbiotic donor microbiomes (see Sup Table 2 for *p*-values).

promoted or inhibited by the same taxa outside the module) and have a common response to external perturbations (e.g., antibiotics). Interaction modules are motivated by both empirical observations that groups of microbial taxa covary[29,30] and theoretical ecology concepts such as guilds, or groups of taxa that utilize resources in a similar way[31]. Modular structure reduces the complexity of the system to be analyzed, which has the potential to increase human interpretability[32,33]. Further, this reduction in complexity enables scalability of the standard gLV model: the number of parameters in the model is reduced from order quadratic in the number of taxa (e.g., all potential pairwise interactions between taxa) to order quadratic in the number of modules (which scales logarithmically with the number of taxa). MDSINE2 additionally extends the gLV model to include stochastic effects and measurement noise error models, enabling full propagation of uncertainty throughout the model and associated confidence measures (e.g., Bayes Factors). We developed and implemented an efficient custom Markov Chain Monte Carlo (MCMC) sampling algorithm to approximate the posterior probability of the model. See Methods and Supplemental Information for complete details on the model and inference algorithm.

Cohorts of gnotobiotic mice "humanized" with fecal transplants from healthy or dysbiotic donors exhibit differential diversity and varied responses to perturbations

To investigate intrinsic differences between dynamics of healthy and dysbiotic microbiomes, we performed experiments with two cohorts of gnotobiotic mice, each cohort gavaged with fecal microbiomes derived from either a healthy donor (H-cohort, n=4) or a dysbiotic donor with ulcerative colitis (D-cohort, n=5) (Figure 2A). After an equilibration period of three weeks, mice were subjected to a sequence of three perturbations (high fat diet (HFD), vancomycin, and gentamicin) designed to provide rich data for dynamical systems inference, by differentially perturbating components of the microbiome (e.g., high fat/simple carbohydrate vs. complex carbohydrate utilizers and bacteria susceptible or resistant to different antibiotics). Mice were separately housed and fecal samples were collected over the 65-day duration of the experiment, with an average of 77 samples per mouse. Samples were interrogated for relative abundance via 16S rRNA amplicon sequencing and total bacterial concentration via qPCR using a universal 16S rDNA primer. The resulting 59 million sequencing reads were bioinformatically processed using DADA2[34] resulting in 1473 Amplicon Sequence Variants (ASVs) that were then further agglomerated into 1224 Operational Taxonomic Units (OTUs) at a 99% sequence similarity threshold. We performed this additional agglomerative step because we observed DADA2-induced artifacts of ASVs dropping in/out throughout time-series (see Sup Information §4 for further details).

Microbiomes between the two cohorts were significantly different both in terms of ecological diversity (Sup Fig 1) as well as in taxonomic composition (Fig 2 and Sup Fig 2). Over the course of the study, the H-cohort had significantly higher alpha diversity than the D-cohort ($p<0.05$ for 56 of the 77 time points, Wilcoxon rank-sum, Sup Fig 1A and Sup Table 1). Beta diversity was significantly different between the two cohorts over the entirety of the time series ($p=0.001$, PERMANOVA, Sup Fig 1B,C). Fold-change analyses showed that taxonomic composition of the microbiomes of mice from the two cohorts also differed significantly. Considering taxonomic composition at steady-state, at the phylum level (Sup Fig 2), abundances of Bacteroidetes and Verrucomicrobia were significantly higher in the H-cohort (see Sup Table 2 for *p*-values). At the family level (Fig 2E, Sup Table 2), seventeen taxonomic groups were significantly different between the two cohorts including higher abundances of Bacteroidaceae, Porphyromonadaceae, Sutterellaceae, Lachnospiraceae, Veillonellaceae, and Akkermansiaceae and lower abundances of Odoribacteraceae, Prevotellaceae, Desulfovibrionaceae in the H-cohort relative to the D-cohort.

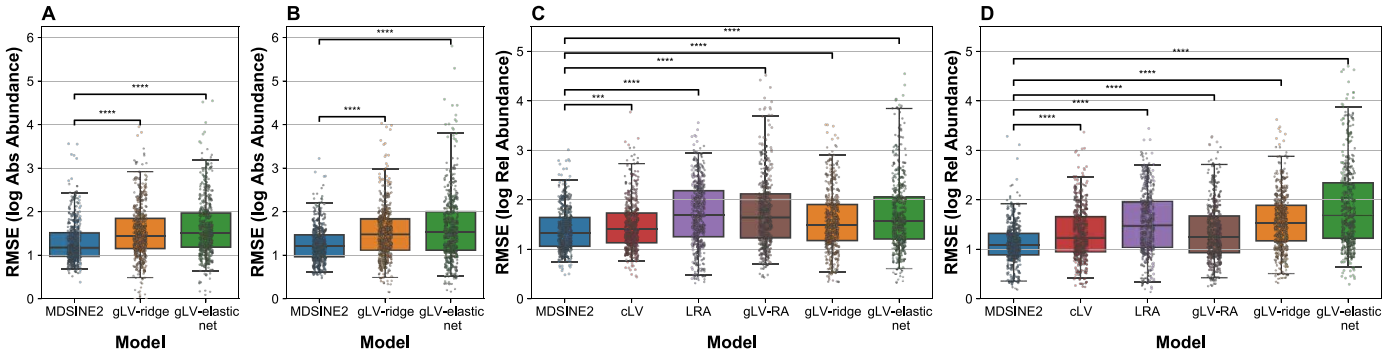

Figure 3: MDSINE2 outperforms state-of-the-art methods in predicting relative and absolute microbial abundances. Performance of methods was evaluated using Root Mean Square Error (RMSE) on forecasted trajectories with hold-one-subject-out cross-validation. MDSINE2, generalized Lotka-Volterra (gLV) with ridge (gLV-ridge) or elastic net regularization (gLV-elastic) forecast absolute abundances; the compositional gLV (cLV), linear dynamics relative abundance (LRA) and gLV trained on relative abundance (gLV-RA) forecast relative abundances. For comparisons in relative abundance space, forecasts of MDSINE2, gLV-ridge/elastic were re-normalized. **(A) and (C)** Comparison of RMSE per OTU forecast for the healthy donor microbiome for absolute and relative abundance, respectively. **(B), (D)** Corresponding results for the dysbiotic microbiome. (****: p<0.0001, ***: p<0.001); one-tailed Wilcoxon signed rank test followed by Benjamini-Hochberg correction for multiple comparisons.

Fold-change analyses showed that responses to the perturbations were generally consistent across the two cohorts at the phylum level (Sup Fig 3A, Sup Table 2), with decreases of Bacteroidetes and increases of Firmicutes and Verrucomicrobia on the high fat diet; decreases of Firmicutes and increases of Proteobacteria on vancomycin; and increases in Firmicutes and decreases in Verrucomicrobia on gentamicin. At the family level (Sup Fig 3B, Sup Table 2), fold-change analyses also found common responses between the cohorts for some bacterial groups, such as decreases in Bacteroidaceae, Prevotellaceae, Sutterellaceae and increases in Peptostreptococcaceae, Desulfovibrionaceae, Enterobacteriaceae Erysipelotrichaceae, Streptococcaceae on the high fat diet; decreases in Lachnospiraceae, Ruminococcaceae and increases in Erysipelotrichaceae on vancomycin; and decreases in Akkermansiaceae on gentamicin. However, the underlying differences in taxonomic composition between the two cohorts and the inability of fold-change analyses to disentangle direct and indirect effects of perturbations (i.e., whether the perturbation itself caused the change or it arose indirectly through chains of microbe-microbe interactions) limits the utility of such analyses for comparing dynamics of ecosystems. MDSINE2 addresses these challenges, as described below, to provide insights into common and differential dynamics between the healthy and dysbiotic donor microbiomes.

MDSINE2 outperforms state-of-the-art methods in forecasting microbial dynamics

We evaluated MDSINE2's ability to forecast held-out microbiome trajectories, an established performance measure[19], against state-of-the-art methods. Previous comparisons have been performed on gnotobiotic datasets with <20 taxa; our dataset is the largest dense time-series with perturbations, to our knowledge, and thus represents a more challenging and realistic benchmark than previously investigated. Consistent with prior work[35], to ensure sufficient time-series information for dynamical systems inference, taxa that were not present at ≥0.01% relative abundance for seven consecutive time-points in at least two mice in each cohort were filtered out. For the H- and D-cohort this resulted in 94 and 113 OTUs respectively, with 60 of those OTUs in common (Sup Figure 4).

We evaluated popular comparator methods that use total bacterial concentration measurements (gLV with ridge regression-based inference [gLV Ridge][17] and gLV with elastic net regularization [gLV Elastic]) and methods that use only relative abundance compositional information (compositional Lotka Volterra [cLV] [20], linear dynamics trained on relative abundances [L-RA], and gLV Elastic trained on relative abundances [gLV-RA]). For forecasting comparisons, we employed one-subject-hold-out training and testing methodology, e.g., holding out all data from one mouse, training on the remaining data, and forecasting all taxa trajectories for the entire 65-day time-series for the held-out mouse given only an initial data point. We evaluated performance using root-mean-squared error over the time-series, e.g., a measure of the difference between the predicted and ground-truth measurement.

MDSINE2 consistently outperformed all the comparator methods (Figure 3). To gain further insight into this result, we analyzed predictive error for each OTU, sorted by its abundance. All the methods performed well on the top 10% of highly abundant OTUs (>10^9 CFU/g) when trained on absolute abundances. But, interestingly, MDSINE2 significantly outperformed the other methods on ~80% of the other taxa, which are at lower abundances (Sup Fig 5, *p*-values in Sup Table 3). MDSINE2's ability to better model a much broader and more diverse set of taxa than the other methods is relevant, because lower abundance taxa have been shown to play important roles in the gut, e.g., *Clostridium scindens*[36] or *Paraclostridium bifermentans*[37] providing resistance to *Clostridioides difficile* infection. Of note, there was no significant difference in performance between MDSINE2 and the comparator methods for the lowest abundance OTUs (bottom ~10% in abundance). This result could be due to a variety of factors, including inconsistent presence of these OTUs in the data (e.g., falling below the limits of detection for some time-points) and insufficient low abundance data for calibrating MDSINE2's error model.

MDSINE2 automatically organizes OTUs into interaction modules that demonstrate coherence across microbiomes

To gain insights into the overall structures of the healthy and dysbiotic microbiome dynamical systems models inferred by MDSINE2, we analyzed interaction modules (Figure 4; 16 modules for the H-cohort, and 11 for the D-cohort), which provide an interpretable organization of the OTUs into groups with common interactions and responses to perturbations.

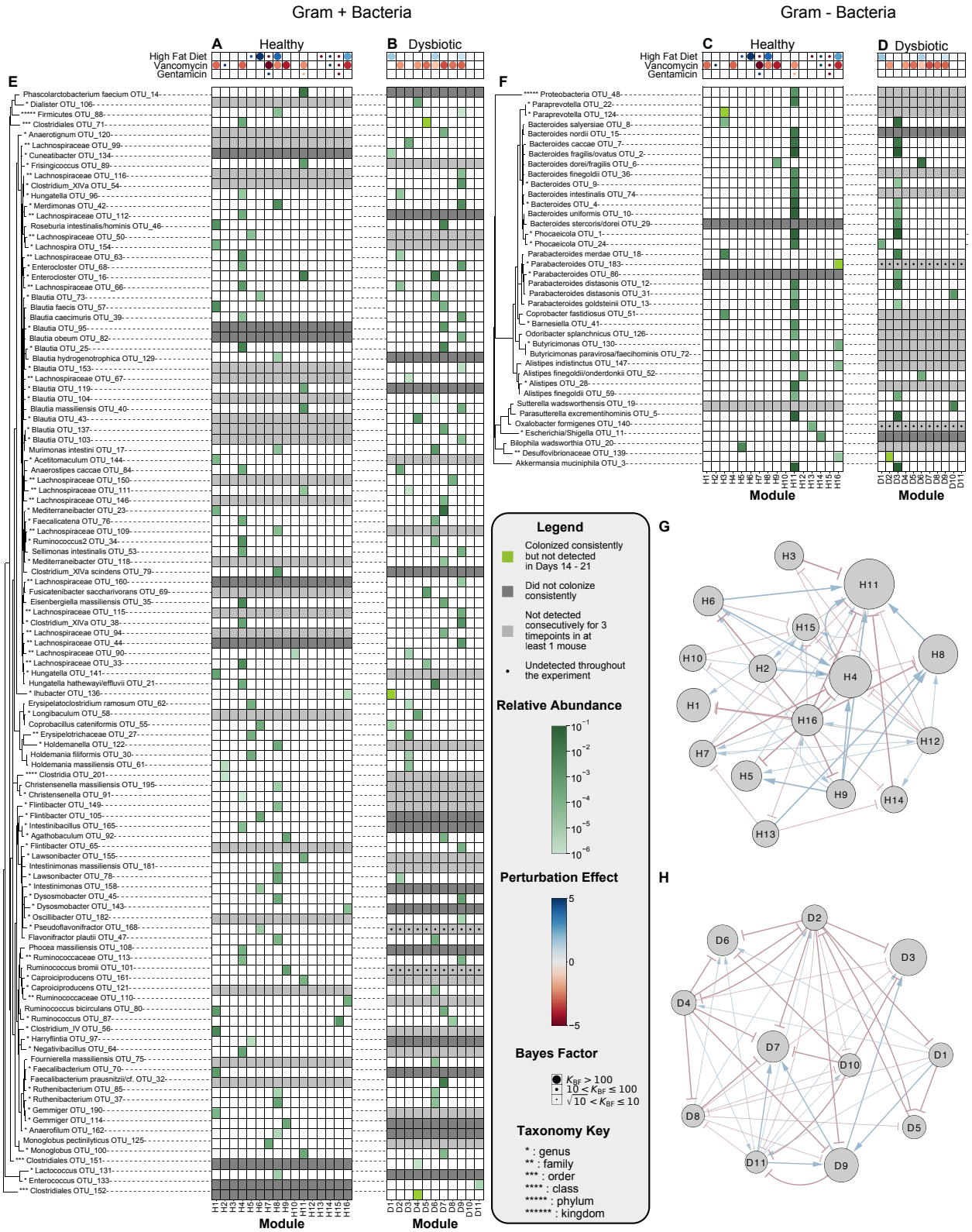

Figure 4: MDSINE2 infers modular representations of complex microbiome dynamics. Our method automatically learns modules of OTUs based on similarity of their dynamic interactions and responses to perturbations. Results are split into Gram-positive and Gram-negative OTUs for display purposes. **(A)** and **(B)** Perturbation effects on Gram-positive OTUs in healthy and dysbiotic microbiomes, respectively. **(C)** and **(D)** Corresponding perturbation effects on Gram-negative OTUs. **(E)** Module memberships for Gram-positive OTUs. Intensity of color in the grid indicates abundance post-colonization and prior to perturbations (average over Days 14 to Day 21). **(F)** Corresponding module memberships for Gram-negative OTUs. **(G)** and **(H)** Inferred module interaction networks for health and dysbiotic microbiomes, respectively. Thickness of edges denotes the strength of evidence; only edges with Bayes Factors > 10 are shown. Sizes of nodes denote the number of OTUs in the module.

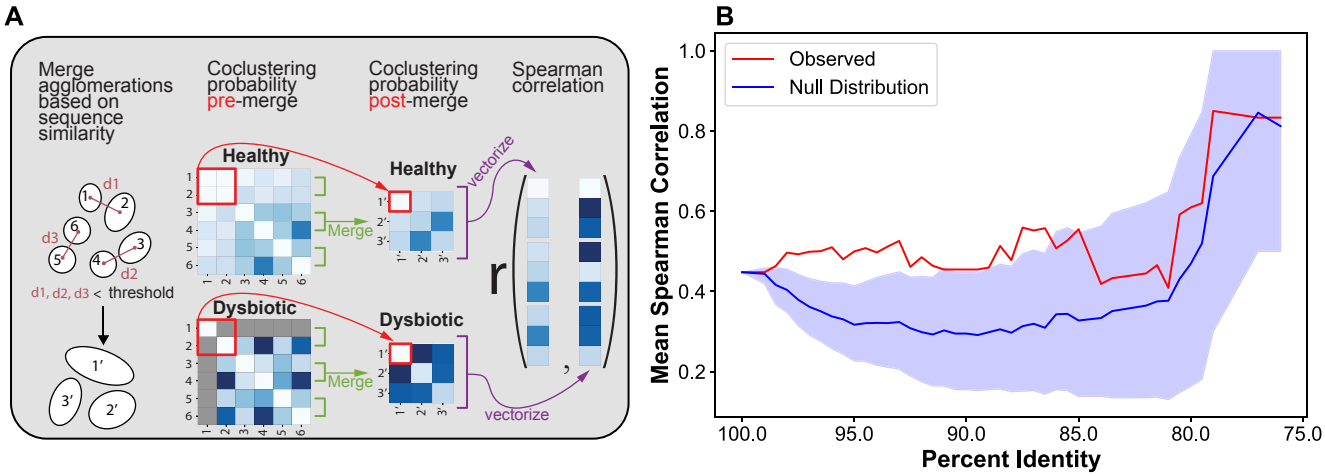

Figure 5: Interaction module membership demonstrates significant consistency across healthy and dysbiotic microbiomes. Phylogenetic Neighborhood Analysis (PNA) allows comparison between complex microbiomes that do not overlap extensively at the OTU level, by comparing successively phylogenetically coarsened representations of the ecosystems. **(A)** Illustration of the PNA algorithm. **(B)** Interaction module membership was significantly correlated between the healthy and dysbiotic microbiomes at ~97% (approximately species taxonomic level) down to ~87% (approximately family taxonomic level) sequence identity, as compared to a null distribution obtained by randomly permuting OTU labels. As expected, at very coarse levels, modular structure is lost (eventually one module is formed) in both the observed and null distributions.

Taxonomy provides one means of interpreting the microbiological context of modules. We found that multiple modules in both the healthy and dysbiotic donor microbes were significantly enriched at the family, order, class, and phylum levels, including expected segregation into the major Gram positive and negative phyla, Bacteroidetes and Firmicutes (Supp Fig 6). As noted above, fold-change analyses cannot distinguish direct from indirect effects. In contrast, MDSINE2, through its underlying dynamical systems model and Bayesian framework, provides quantitative estimates of the evidence for direct effects via Bayes Factors (BF)[38]. Thus, we investigated whether the direct effects of perturbations on taxa corresponded to known microbiology established in previous studies. The perturbation that had the largest impact was vancomycin with *decisive evidence* (BF>100) for direct effect on seven modules in the H-cohort and also seven modules in the D-cohort, with negative signs (repressive effects) for all the modules. In 86% of the affected modules in the H-cohort and 100% in the D-cohort, the majority of the taxa were Firmicutes, corresponding to expected vancomycin susceptibility of members of this phylum. MDSINE2 identified the high-fat diet as the next strongest perturbation, with *decisive evidence* for direct effects on three modules within each cohort (positive effect in all cases). In 66% of affected modules in the H-cohort and 100% in the D-cohort, the majority of taxa were again Firmicutes, consistent with prior studies that have shown greater capacity for members of this phylum to grow on fat sources[39]. No modules showed decisive evidence for the gentamicin perturbation; this may be because gentamicin's primary activity is against aerobic, Gram-negative bacteria (e.g., *Enterobacteriaceae*), which were small components of the gut microbiomes studied. Although these results indicate that many modules captured taxonomic relationships, taxonomy does not reflect the detailed evolutionary signals encapsulated in phylogenic analyses, and is thus a fairly blunt instrument for interpreting microbiological context.

As another means to assess the quality and utility of modules as an organizing principle for gut microbiome dynamics, we investigated their coherence across different microbiomes in terms of phylogeny, i.e., whether taxa in a module in one microbiome tend to be in a module in another microbiome. However, direct comparison of OTU membership in interaction modules across microbiomes is challenging, because of limited overlap in OTUs (in our dataset, 60 OTUs are common, out of 94 in the H- and 113 in the D-cohort). To address this challenge, we used an approach we term Phylogenetic Neighborhood Analysis (Figure 5). Briefly, the method merges OTUs into successively phylogenetically coarser sets and then evaluates module coherence of the resulting sets between microbiomes. As expected, module coherence was not significantly different from chance at high phylogenetic similarity (corresponding to low overlap of species present in both microbiomes) or at low phylogenetic similarity (corresponding to highly heterogeneous groups of taxa that would not be expected to have similar behavior). However, module coherence between the H- and D-cohort microbiomes was significant between 86% and 95% sequence identity (approximately corresponding to genus to family level in terms of taxonomy). Coherence at this level of sequence identity, where we may expect to see similar phenotypes and niche behaviors of microbes, suggests that the modules capture relevant microbiological relationships.

The dysbiotic microbiome exhibits intrinsic instability driven by competitive cycles

To evaluate intrinsic stability behavior relevant to microbial ecosystems, we performed *in silico* "knock-down" experiments designed to mimic strong antibiotic treatments, testing systems' robustness to perturbations, or ability to return to steady-state after disturbances. Briefly, random sets of bacteria (ranging from 10% to 70% of total taxa) were repeatedly sampled and a negative perturbation was applied to growth rates of the bacteria. We then compared post-perturbation steady-states to the steady-states of the unperturbed ecosystems (Fig 6A), and also evaluated the ecological diversities of perturbed ecosystems (Figure 6B). The D-cohort demonstrated a significantly larger deviation of the post-perturbation steady state when compared to the H-cohort, (*p*-value below machine precision, Mann-Whitney U, Sup Table 5). Further,

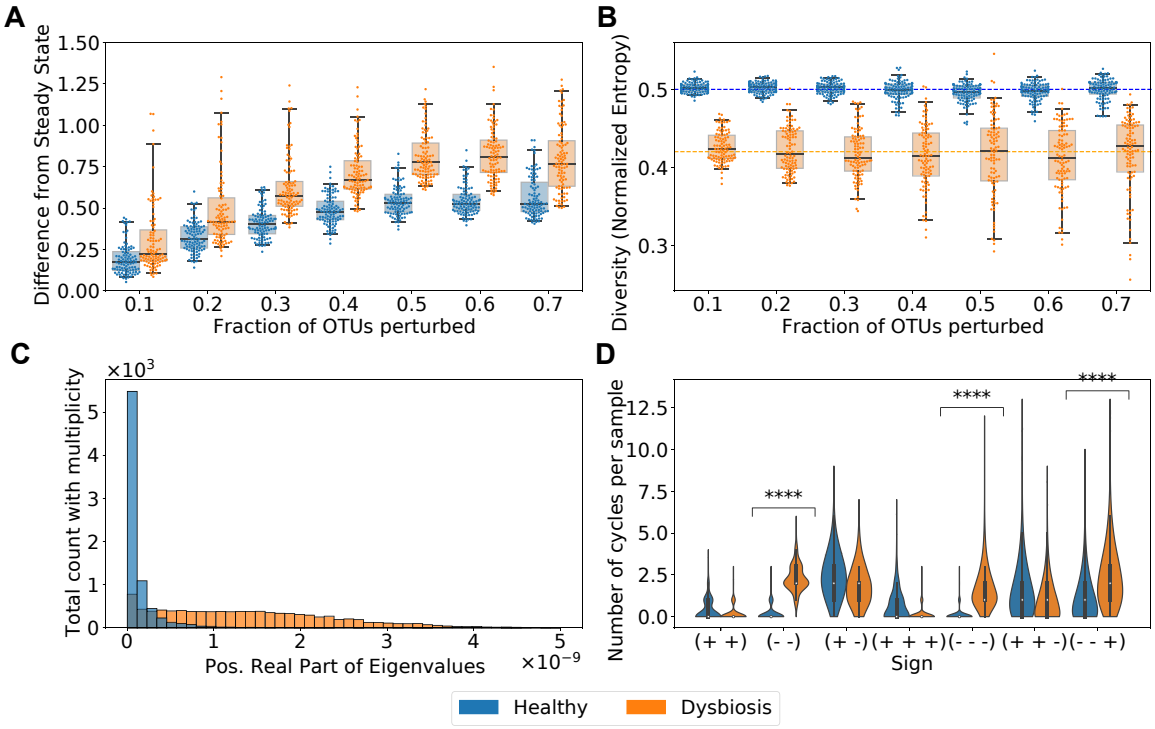

Figure 6: The dysbiotic microbiome exhibits instability as assessed by *in silico* simulations and formal analyses. **(A)** The dysbiotic microbiome exhibits greater deviations from baseline and more variability in deviations, when subjected to strong *in silico* perturbations (i.e., analogous to antibiotic exposures). Fractions of OTUs were sampled repeatedly and negatively perturbed to simulate an effect analogous to an antibiotic disturbance to the ecosystem. **(B)** The ecological (alpha) diversity of the dysbiotic microbiome also shows post-perturbation decreases and greater volatility, when the systems are subjected to the same perturbations as in (A). The differences were significant for all fractions tested in both (A) and (B) (see Supp Table 5 for *p*-values). **(C)** Analytical assessment of stability demonstrated lower stability of the dysbiotic microbiome (extreme skewing of the dysbiotic system's eigenvalues to right half of the complex plane with greater positive real parts). **(D)** Topological analysis revealed significantly more competitive cycles in the dysbiotic microbiome (see Supp Table 5 for *p*-values), which is also a hallmark of instability.

credible intervals of deviations were larger for the D-cohort, demonstrating greater variability in post-perturbation steady states, which is an additional hallmark of instability. Ecological diversity analyses similarly showed strikingly different behaviors of the H- and D-cohorts: even when large fractions of taxa were perturbed, the H-cohort microbiome tended to return to pre-perturbation diversity levels with high probability, whereas the D-cohort microbiome exhibited increasing probability of dramatic collapses in diversity as more taxa were perturbed (Fig 6B). These results indicate an intrinsic inability of the dysbiotic microbiome to recover consistently from strong perturbations, such as could occur in the setting of antibiotic administration, infection, or other major insults to the ecosystem.

We additionally evaluated the stability of the two ecosystems using mathematical criteria from dynamical systems theory. These criteria require that the eigenvalues of the matrix describing the pairwise interactions from the gLV dynamics equations have negative real parts.[40] Because MDSINE2 estimates the full posterior probability distribution of inferred models, we can compute a readily interpretable measure of stability for stochastic microbial ecosystem dynamics: the probability of the system having all negative eigenvalues (see Sup Information §5). For this measure, we found that the H-cohort microbiome has a 31% higher probability of being stable than the D-cohort microbiome (Fig 6C), providing additional evidence for intrinsic instability of the dysbiotic microbiome.

We next sought to identify features of the ecological network inferred by MDSINE2 that could explain the differences in stability seen between the H- and D-cohort microbiomes. The overall edge density differs between the two networks (22% for H- and 40% for D-cohort) as well as the ratios of positive to negative interactions (1:1 for H- and 1:2 for D-cohort). The higher edge density alone suggests that the D-cohort dynamical system is potentially less stable[9,40], due to a higher probability of containing *feedback cycles* in the network. Stability and control theory have established that the feedback cycle is the core topological feature driving stability[41]. Pairwise interactions, the simplest form of feedback cycles, have particular interpretations in ecology, and their contributions to stability are well-characterized for linear and gLV dynamical systems[10]: mutualism (+/+) and competition (−/−) are destabilizing, and parasitism (+/−) is stabilizing (Sup Fig 7). For length three cycles and higher, more complex ecological interactions arise, and any sign combination is potentially destabilizing[42]. To investigate these topological features in the H- and D-cohort dynamical systems, we tested for the enrichment of length two and three cycles (Fig 6D). The most differentially abundant two-cycle motif was competition (-/-), with significantly more such cycles present in the D-cohort than the H-cohort (*p*-value below machine precision, Mann-Whitney U). The most differentially abundant cycle of length three was also an all-competition cycle, again with significantly more of those cycles in the D-cohort than the H-cohort (*p*-value below machine precision, Mann-Whitney U). These findings suggest that competitive cycles drive the intrinsic instability observed in the dysbiotic microbiome dynamical system.

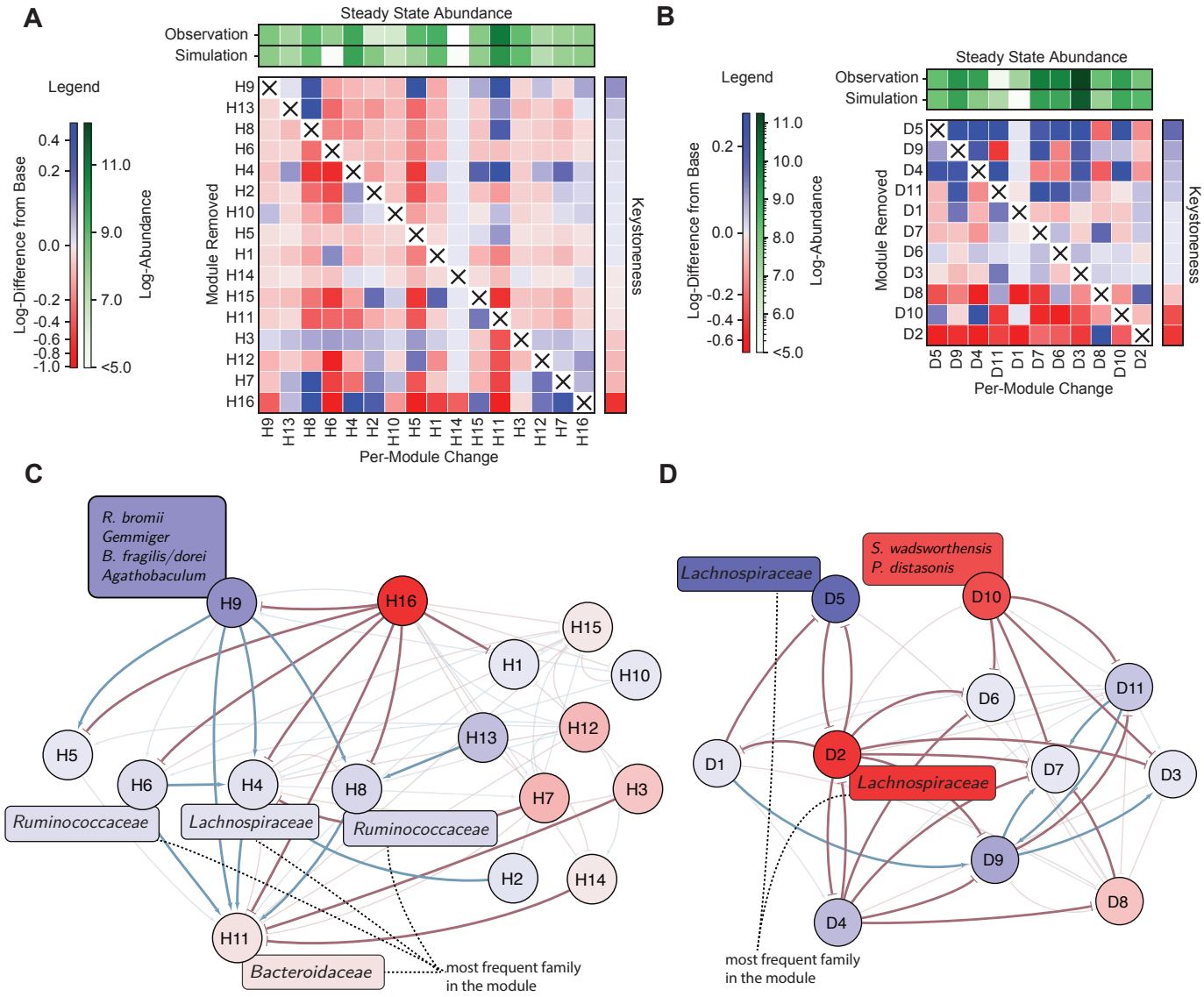

Figure 7: Keystoneness analysis characterized the relative importance of each module for maintaining microbiome states and revealed chains of ecological interactions governing dynamics. (A) and **(B)** Keystoneness maps for healthy and dysbiotic donor microbiomes provide quantitative rankings of the effect of removing each module on the remainder of the ecosystem. **(C)** and **(D)** By tracing interactions from the top positive and negative keystoneness modules, chains of ecological interactions are highlighted. In the healthy donor microbiome, chains of positive ecological interactions were seen, with the head node (H9) consisting of taxa with strong starch degradation capabilities. In contrast, the dysbiotic donor microbiome is dominated by competitive interactions. Thick edges are bayes factor greater than 10 (strong evidence) and thin transparent edges are Bayes factors between $\sqrt{10}$ and 10 (substantial evidence).

Keystoneness analysis reveals modules of bacteria with functional ecological roles that promote stability or instability in healthy or dysbiotic microbiomes

To evaluate quantitatively the relative importance of interaction modules in the ecosystem, we performed a module keystoneness analysis (Fig 7A,B). Modules with the highest positive or negative keystoneness values characterize groups of bacteria that are most critical to maintaining the structure of the ecosystems. Positive keystone modules ("promoters") are those that when removed result in a reduction in the microbial abundances of the other community members; negative keystone modules ("suppressors") are those that when removed result in increases of abundances of the other community members.

For the H-cohort, the top positive and negative keystoneness modules are H9 and H16 (Figure 7C), respectively. These modules act as *hubs* (≥4 outgoing edges) in the network, with H9 promoting all the modules it interacts with, and H16 suppressing all the modules it interacts with, respectively, suggesting opposite ecological roles for H9 and H16. Interestingly H9 and H16 participate in a regulatory feedback loop as well, a topology that provides robustness to external disturbances. Interestingly, all the members of H9 (OTU6 *Bacteroides fragilis/dorei*, OTU92 *Agathobaculum*, OTU101 *Ruminococcus bromii* and OTU114 *Gemmiger*) have the capability to degrade starches. *R. bromii* has previously been identified as a keystone species for its ability to degrade resistant starches RS(3), with preference for large α(1-4)-linked oligosaccharides. OTU114 *Gemmiger* can metabolize resistant

starches RS(2)[43], and this genus has been shown to be absent or at reduced abundance in patients with inflammatory bowel diseases[44,45]. The other modules in the positive interaction chain with H9 include H4, H5, H6, H8, and H11 (H4, H5, H6, and H8 have directed edges promoting H11 as well). Most of the taxa in H4, H6, and H8 are in the Order *Clostridiales*, with H4 taxa more predominantly from the Family *Lachnospiraceae*, and H6 and H8 taxa more predominantly from the Family *Ruminococcaceae*. Module H5 consists of predominantly taxa in the Order *Erysipelotrichales*. Finally, H11 is the largest module in the H-cohort ecosystem, and contains a majority of the Gram-negative bacteria in that microbiome. Of note, the topology identified by MDSINE2, as well as the metabolic products of taxa in these modules suggest a possible cross-feeding network beginning with primary starch degradation in H9, followed by fermentation in H4, H5, H6, and H8, producing acetate, lactate, and simple carbohydrates.

For the D-cohort, the top positive and negative keystone modules were D5 and D2 (Figure 7B), respectively, with all the OTUs in D5 belonging to the Family *Lachnospiraceae* and all but one of the OTUs in D2 also belonging to the Family *Lachnospiraceae*. When one considers the complete network, both modules effectively have suppressive activity: D5 is in competition with D2, which is a hub suppressing all its neighbors (Fig 7D). Thus, D5 indirectly promotes the growth of other taxa in the cohort (by suppressing a suppressor), and in so doing also participates in a potentially unstable feedback cycle with D2. Another negative keystone module is D10, a directly suppressive hub like D2, which contains two taxa: OTU19 *Sutterella wadsworthensis* and OTU31 *Parabacteroides distasonis*. Interestingly, *S. wadsworthensis* has previously been positively associated with inflammatory bowel disease in observational studies and was found to be a negatively associated with response to fecal microbiota transplant (FMT) in UC patients[46-49]. *S. wadsworthensis* has several properties that give it potential competitive advantages in the gut, particularly in dysbiotic environments, including being a facultative anaerobe, bile resistant, a nitrate reducer, capable of degrading host IgA, and able to adhere to mucus and extracellular protein matrices. In contrast to the H-cohort where the top negative keystoneness module is in a stable feedback loop with another module, the D-cohort's top negative keystoneness module can potentially grow unchecked, suppressing other community modules, providing further insights into how the dysbiotic microbiome may be destabilized.

Discussion

A important innovation of our method is the ability to analyze microbial dynamical systems in terms of modules. Modularity has been extensively exploited for understanding other complex biological systems, such as mammalian genetic regulatory networks[50-52], but relatively underexploited in the microbiome field[29,30]. The conventional approach in the microbiome field is to use taxonomy to define groups for analysis. One shortcoming of taxonomic approaches is that the names of many bacteria are historical artifacts that do not accurately represent evolutionary relationships. Another significant issue is that taxonomic groupings do not necessarily correspond to any clear functions or behaviors in an ecosystem. In contrast, our approach automatically learns groupings of microbes based on shared interaction structures in the dynamic ecosystem and responses to external perturbations, both of which are functional properties of the microbes. Indeed, our method identified groups of bacteria that are quite taxonomically and phylogenetically distinct, yet share a common function, such as the taxa in the top positive keystone module in the healthy microbiome, consisting of *Ruminococcus*, *Bacteroides*, *Agathobaculum*, and *Gemmiger* that share a common ability to degrade starches.

Our results also demonstrate the utility of analyzing intrinsic dynamical systems properties of complex microbiomes. Reductionist approaches seeking to find pathobiont species to explain dysbiotic states in human microbiomes, such as in ulcerative colitis, have not yielded clear results, despite intensive efforts. In some ways, this is not surprising: the human microbiome constitutes a highly complex and dynamic ecosystem. Our results suggest that dysbiotic behavior of these microbiomes can be explained at the *systems level*, in terms of interactions in the underlying ecological network. We found that microbiomes, when taken out of the human hosts and transplanted into genetically identical mice that did not have disease symptoms, still exhibited intrinsic differences in stability driven by the topology of interactions. This topology is not discoverable through standard differential abundance analyses of cross-sectional studies. Such analyses return lists of microbes that differ between cohorts but offer no further means to understand the reasons behind these differences. In contrast, dynamical systems analyses offer insights into how ecosystems behave, including under physiologically relevant perturbations. For example, MDSINE2 identified a key network of interactions in the healthy microbiome, which could constitute a cross-feeding network that begins with organisms that degrade resistant starches and ultimately result in production of butyrate, a primary energy source of intestinal cells important in maintaining proper intestinal barrier function[22,53,54]. Restoration of taxa in this chain in dysbiotic microbiomes might therefore increase stability, and moreover produce direct benefits to the host.

There are several directions for future work on the computational model. First, MDSINE2's model of measurement noise, a well-established model based on the negative binomial distribution, could be extended to include zero inflation or nonparametric features (which could be particularly advantageous for metagenomics or other data types that may not fit the negative binomial model well). A second direction for extension is the gLV model, which only allows for pairwise, "mass action" style interactions. Although it would be conceptually straightforward to include higher-order interactions, it remains unclear how prevalent such interactions are in mammalian microbiomes[55]; another approach would be to include interactions with saturation behavior, or even learn nonparametric models for interactions to avoid needing to prespecify their forms. A third, and particularly exciting area for extension, would be incorporating additional data modalities to more fully capture the host-microbial ecosystem, such as metagenomic, metatranscriptomic and metabolomic data sources.

Conclusion

We have introduced MDSINE2, a computational method for accurately inferring interpretable dynamical systems models of the microbiome at scale, and demonstrated on a new densely-sampled microbiome time-series dataset from "humanized" gnotobiotic mice that our approach outperforms other methods when forecasting microbiome dynamics. Moreover, our approach identified intrinsic instability in microbial dysbiosis, driven by competitive cycles and the absence of chains of stabilizing interactions found in the healthy microbiome. Our approach provides new tools for characterizing the dynamical systems behaviors of complex host-microbial ecosystems and holds promise for guiding rational design of interventions to stably alter human microbiomes for prophylactic or therapeutic purposes.

Methods

MDSINE2 Model

Overview

Our statistical model of microbial dynamics is a fully Bayesian hierarchical model based on continuous-time stochastic generalized Lotka-Volterra (gLV) dynamics:

$$\mathrm{d}\boldsymbol{x}_{s,i} = \boldsymbol{x}_{s,i}(t)\left[\left(1+\sum_{p=1}^{P}\boldsymbol{\gamma}_{p,\boldsymbol{c}_i}\boldsymbol{z}^{(\gamma)}_{p,\boldsymbol{c}_i}h_p(t)\right)\boldsymbol{a}_{1,i} - \boldsymbol{a}_{2,i}\boldsymbol{x}_{s,i}(t) + \sum_{j:\boldsymbol{c}_j\neq\boldsymbol{c}_i}\boldsymbol{b}_{\boldsymbol{c}_i,\boldsymbol{c}_j}\boldsymbol{z}^{(b)}_{\boldsymbol{c}_i,\boldsymbol{c}_j}\boldsymbol{x}_{sj}(t)\right] + \boldsymbol{x}_{s,i}(t)\mathrm{d}\boldsymbol{w}_{s,i}. \quad (1)$$

This formulation of stochastic behavior models multiplicative random effects on microbial abundances, which could arise from a variety of phenomena, such as temporal host, environmental or dietary fluctuations that result in short time-scale increases or decreases in abundance of each taxa.

The abundance of taxa i in time-series s (e.g., biological replicate) is denoted as $\boldsymbol{x}_{si}$. MDSINE2 probabilistically assigns each taxa to an interaction module, where $\boldsymbol{c}_i$ denotes the module assignment for taxa i. The growth rate and self-interaction random variable for taxa i are denoted $\boldsymbol{a}_{1,i}$and $\boldsymbol{a}_{2,i}$, respectively. The P external perturbations are accounted for by the random variables $\boldsymbol{\gamma}_{p,\boldsymbol{c}_i}$ that denote the effect of perturbation p on taxa i's growth rate; $\boldsymbol{z}^{(\gamma)}_{p,\boldsymbol{c}_i}$ is a corresponding random indicator variable that probabilistically selects whether the perturbation affects the interaction module. The function h_p has a value of 1 during the time-period when the p-th perturbation is active and a value of 0 otherwise. The strength of the microbial interaction from taxa j to taxa i is denoted $\boldsymbol{b}_{\boldsymbol{c}_i,\boldsymbol{c}_j}$, with $\boldsymbol{z}^{(b)}_{\boldsymbol{c}_i,\boldsymbol{c}_j}$ the corresponding random indicator variable for that microbial interaction. The stochastic variation of the microbial abundances over time is captured by the variable $\boldsymbol{w}_{s,i}$, specifying geometric Brownian motion for the stochastic component (e.g., a multiplicative stochastic process on the microbial abundance).

To perform inference, we perform a first-order discretization (see Supplemental Information) to obtain the discrete-time latent trajectories:

$$\log\left(\boldsymbol{x}_{s,i}(k+1)\right) \sim \mathrm{Normal}\left(\log\left(\mu_{s,i}(k+1)\right), \Delta_{s,k}\boldsymbol{\sigma}^2_w\right) \quad (2)$$

where

$$\log\left(\mu_{s,i}(k+1)\right) = \log\left(\boldsymbol{x}_{s,i}(k)\right) + \Delta_{s,k}\left[\boldsymbol{a}_{1,i}\left(\boldsymbol{1}+\sum_{p=1}^{P}\boldsymbol{\gamma}_{\boldsymbol{c}_i,\boldsymbol{p}}\boldsymbol{z}^{(\gamma)}_{\boldsymbol{c}_i,\boldsymbol{p}}h_p(k)\right) - \boldsymbol{a}_{2,i}\boldsymbol{x}_{s,i}(k) + \sum_{j:\,\boldsymbol{c}_j\neq\boldsymbol{c}_i}\boldsymbol{b}_{\boldsymbol{c}_i,\boldsymbol{c}_j}\boldsymbol{z}^{(\boldsymbol{b})}_{\boldsymbol{c}_i,\boldsymbol{c}_j}\boldsymbol{x}_{s,j}(k)\right]$$

Here, $\Delta_{s,k} = t_{s,k+1} - t_{s,k}$, the difference between adjacent time-points for the same time-series s.

Given relative abundance time-series data (e.g., 16S rRNA amplicon sequencing) and measurements of microbial concentrations (e.g., 16S rRNA amplicon qPCR with universal primers), MDSINE2 uses a custom Markov Chain Monte Carlo (MCMC) algorithm to infer the full posterior probability distribution. Below we give additional details on the model, including prior probability distributions on variables; for complete mathematical and algorithmic details, see Supplemental Information.

Interaction Modules

We employ a Dirichlet Process (DP) prior[56] to model interaction modules. The expected number of modules under this prior probability distribution is $\approx \alpha\log\frac{N+\alpha}{\alpha}$, where N is the number of taxa and α is the concentration parameter[57]. This property is desirable for scaling to large ecosystems, as the expected number of microbial interactions in our model scales as $O(\log(N)^2)$ (as opposed to $O(N^2)$ in the standard gLV model). We place a diffuse Gamma prior on the concentration parameter as described in[58]. Our formulation allows us to marginalize out the interaction and perturbation parameters during inference, which greatly increases efficiency[56]. See Supplemental Information for complete details.

Interaction Parameters and Perturbation Effects

To facilitate modularity and interpretability of inferred interaction networks, we assume no intra-module interactions and model only inter-module interactions, $\boldsymbol{b}_{\boldsymbol{c}_i\boldsymbol{c}_j}$. We assume perturbations (e.g., antibiotics or dietary changes) have module-specific effects, $\boldsymbol{\gamma}_{\boldsymbol{c}_i}$. Further, we model the presence/absence of module-module interactions and module-perturbation effects by using the binary indicator variables $\boldsymbol{z}^{(\boldsymbol{b})}$ and $\boldsymbol{z}^{(\boldsymbol{\gamma})}$, respectively. These binary indicators allow the model to infer the structural edges that specify the underlying network topology between modules. Additionally, this formulation allows for direct calculation of the statistical evidence for presence of each interaction or perturbation effect using Bayes factors. See Supplemental Information for full details.

Measurement Model

The observed data are sequencing counts $\boldsymbol{y}_{s,i}(k)$ of taxa and qPCR measurements $\boldsymbol{Q}_{s,\mathrm{r}}(k)$ of bacterial concentrations, where *j* indexes the qPCR measurement replicates. Sequencing counts are modeled using a negative binomial distribution[59]

$$\boldsymbol{y}_{s,i}(k) \mid \boldsymbol{x}_{s,i}(k), r_{s,k} \sim \mathrm{NegBin}\left(\varphi\left(\boldsymbol{x}_{s,i}(k), r_{s,k}\right), \epsilon\left(\boldsymbol{x}_{s,i}(k), d_0, d_1\right)\right)$$

$$\varphi\left(\boldsymbol{x}_{s,i}(k), r_{s,k}\right) = r_{s,k}\frac{\boldsymbol{x}_{s,i}(k)}{\sum_j \boldsymbol{x}_{s,j}(k)}$$

$$\epsilon\left(\boldsymbol{x}_{s,i}(k), d_0, d_1\right) = \frac{d_0}{\boldsymbol{x}_{s,i}(k)/\sum_j \boldsymbol{x}_{s,j}(k)} + d_1$$

Here $r_{s,k}$ is the total number of reads for the sample in time-series *s* at time $t_{s,k}$, and d_0 and d_1 parameterize the function $\epsilon(\cdot)$, which specifies the Negative Binomial dispersion parameter. We fit the parameters d_0 and d_1 using data from replicates (see below).

We model the qPCR measurements with a log-normal distribution:

$$\log\left(\boldsymbol{Q}_{s,r}(k)\right) \sim \mathrm{Normal}\left(\log\left(\sum_i \boldsymbol{x}_{s,i}(k)\right), \sigma^2_{\boldsymbol{Q}_s(k)}\right)$$

Here, $\sigma^2_{\boldsymbol{Q}_s(k)}$ is the empirical variance of the set of qPCR measurement replicates for time-series s at time $t_{s,k}$. See Supplemental Information for complete mathematical details of the measurement model and inference procedure.

Software

MDSINE2 was implemented in Python 3.7 using the Numpy[60], Scipy[61], Numba[62], Matplotlib[63], and Seaborn[64] packages. The software is publicly available under the Gnu General Public License v3.0 (https://github.com/gerberlab/MDSINE2). The input to MDSINE2 consists of five tab-delimited files: (1) list of the sequence and taxonomic label for each taxa, (2) table of counts for each taxa in each sample, (3) table specifying the time points at which each sample was collected for each subject, (4) table of qPCR values for each sample, and (5) table of perturbation names, start times, end times, and associated subjects that received the perturbation. The software outputs inference results in two files: (a) a Python pickle file that contains the MDSINE2 inference object used to perform the inference, and (b) a HDF5 file containing all the MCMC posterior samples. Once inference is complete, the software includes functionality to visualize and interpret the posterior samples, including visualizing trajectories, module networks (with a Cytoscape[65] export option) and keystoneness, as well as generating text files with summaries of posterior distributions. See online software documentation for complete details. We also give demos of the functionalities in the binder tutorials. The tutorials (Google Colab) can be accessed in the folder https://github.com/gerberlab/MDSINE2_Paper/tree/master/google_colab.

Gnotobiotic Experiments and Microbiome Data Generation

Mouse Experiments

Two cohorts of male C57Bl/6 germfree mice (n=4 for the H-cohort and n=5 for the D-cohort) were used in the experiments (BWH IACUC: 2016N000141). Mice were singly housed in Optmice cages within the Massachusetts Host-Microbiome Center (MHMC) at Brigham and Women's Hospital.[26] The mice in each group were given a Fecal Microbiota Transplant (FMT) from two de-identified human stool donors (a healthy donor for the H-cohort and a donor with ulcerative colitis for the D-cohort) from an ongoing study at Brigham and Women's Hospital (IRB# 2017P002420). Per the study protocol, samples were flash frozen without cryoprotectants and stored at -80°C. Material for FMTs was prepared by thawing the stool samples and homogenizing in 5 mL of pre-reduced 1x Phosphate Buffered Saline (PBS) with 0.05% cysteine inside an anaerobic chamber. Germfree mice were then orally gavaged with 200µl/mouse of FMT material. Post-gavage, mice were equilibrated for 3 weeks before beginning a series of three perturbations: high fat diet (HFD), vancomycin, and gentamicin (in that order). Each perturbation lasted for one week, followed by a one-week normalization period off perturbations. Aside from the HFD perturbation, mice were maintained on standard MHMC gnotobiotic mouse chow (Autoclavable Mouse Breeder Diet 5021; LabDiet). For the HFD perturbation, Research Diets D12492 (60 kcal% of fat) was used. For the vancomycin perturbation, drinking water was replaced with water containing vancomycin at a concentration of 100 ug/mL and 3% sucralose (filter sterilized). For the gentamicin perturbation, drinking water was replaced with water containing gentamicin at a concentration of 4 ug/mL and 3% sucralose (filter sterilized). In all situations, mice were allowed to eat and drink ad libitum. Mouse fecal pellets were collected in triplicate based on the sample collection timeline detailed in Figure 2. We also obtained additional samples to generate data for fitting the d0 and d1 parameters in our amplicon sequencing measurement noise model. For this purpose, a total of nine fecal pellets, three pellets on each of the three consecutive days (8, 9, 10) were collected from mouse 2. Each fecal pellet was divided into two parts. This resulted in 18 samples that were then processed through the entire sequencing pipeline, from DNA extraction through sequencing. To collect fecal pellets, each mouse was removed from the Optimice cage and placed inside an autoclaved Nalgene cup. After pellets were produced, mice were placed back in their cages and samples were collected from the cup with autoclaved forceps. Samples were placed in cryovial tubes and snap frozen in liquid nitrogen immediately, then stored at -80°C. At the end of experiments, mice were euthanized by overdose on inhaled vapors of isoflurane administered in an anesthesia chamber followed by cervical dislocation. These procedures are in accordance with the recommendations of the Panel on Euthanasia of the American Veterinary Medical Association.

DNA Extraction, 16S rRNA Amplicon Sequencing and qPCR

For DNA extraction, all samples were processed using the standard protocol[66] at the Massachusetts Host-Microbiome Center (MHMC), which uses the Zymo Research ZymoBIOMICS DNA 96-well kit according to manufacturer instructions with the addition of bead beating for 20 minutes. Amplicon sequencing and qPCR were also performed using the standard MHMC protocol. Briefly, for amplicon sequencing, the v4 region of 16S rRNA gene was PCR amplified using 515F and 806R primers[67]: 5'-[Illumina adaptor]-[unique bar code]-[sequencing primer pad]-[linker]-[primer]

- (fwd primer): AATGATACGGCGACCACCGAGATCTACAC-NNNNNNNN-TATGGTAATT-GT-GTGCCAGCMGCCGCGGTAA
- (rev primer): CAAGCAGAAGACGGCATACGAGAT-NNNNNNNN-AGTCAGTCAG-CC-GGACTACHVGGGTWTCTAAT

Following PCR of the v4 region, 250 bp paired end reads were generated on an Illumina MiSeq with the following custom primers with index primer: ATTAGAWACCCBDGTAGTCC-GG-CTGACTGACT.

- 5'-[sequencing primer pad]-[linker]-[primer]
  Read 1: TATGGTAATT-GT-GTGCCAGCMGCCGCGGTAA
- 5'-[primer]-[linker]-[sequencing primer pad]
  Read 2: AGTCAGTCAG-CC-GGACTACHVGGGTWTCTAAT

qPCR for estimating total bacterial concentration was performed using universal 16S primers and a standard curve prepared from dilutions of *Bacteroides fragilis* (ATCC 51477). Samples were loaded into 384 well plates via the Eppendorf EP Motion liquid handler and then run on a QuantStudio 12K Flex Real-Time PCR System (ThermoFisher) using TaqMan Universal Master Mix II no UNG kit (ThermoFisher 4440040), TaqMan Gene Expression Assay (ThermoFisher 4331182), probe set Dye: FAM, Quencher: NFQ-MGB and Reference Dye: Rox for quantification (ThermoFisher assay ID Pa04230899_s1), all according to manufacturer's instructions.

Bioinformatics

Generating ASV tables From Amplicon Reads

We generated an ASV read count table and assigned taxonomy using *DADA2 v1.16* according to the standard pipeline using pseudo-pooling[34]. Forward reads were trimmed to a length of 240 and reverse reads were trimmed to a length of 160. Our function calls for these core steps in the DADA2 pipeline were:

```
out <- filterAndTrim(fnFs, filtFs,
    fnRs, filtRs, truncLen=c(240,160),
    maxN=0, maxEE=c(2,2), truncQ=2,
    rm.phix=TRUE, compress=TRUE,
    multithread=TRUE)
errF <- learnErrors(filtFs,
    multithread=TRUE, randomize=TRUE,
    nbases=1e8, pool = "pseudo")
errR <- learnErrors(filtRs,
    multithread=TRUE, randomize=TRUE,
    nbases=1e8, pool = "pseudo")
dadaFs <- dada(filtFs, err=errF,
    multithread=TRUE, pool = "pseudo")
```

```
dadaRs <- dada(filtRs, err=errR,
    multithread=TRUE, pool = "pseudo")
```

To assign taxonomic labels to ASVs, we used DADA2-formatted reference databases RDP trainset 16 and Silva version 138. When using `assignTaxonomy` in DADA2, we specified the maximum number of multiple species assignments to be 2. For species assignments, if one database returned a species assignment and the other did not, we labelled the ASV with the species from the database that returned the assignment. If both databases returned species assignments, but they were discordant, we set the assignment to the union of the returned assignments. If the total number of possible species assigned was greater than 4, then we did not set a species assignment. This process resulted in 1473 ASVs in total. The DADA2 script used for this study is located in https://github.com/gerberlab/MDSINE2_Paper/tree/master/google_colab

ASV Aggregation into OTUs

We performed agglomerative clustering of ASVs into OTUs (average linkage, using Hamming distance = 3 base pairs). This process resulted in 1224 OTUs in total from the original 1473 ASVs. Consensus sequences
for OTUs were determined by requiring >=65% consensus at each position, otherwise the position was assigned an 'N'. The taxonomic assignment for each OTU was derived as follows. If the taxonomy of all ASVs within an OTU agreed down to the species level, then we marked that OTU with that taxonomic label. If the ASVs instead only agreed down to the genus level with differing species, then we marked that OTU with that genus, and the species of the OTU was labelled with the union of all of the constituents' species. Otherwise, if there were any conflicts at the genus level or above, we assigned the taxonomy using RDP's Naïve Bayes classifier[68] with the OTU's consensus sequence as input.

Phylogenetic Placement of Sequences

We performed phylogenetic placement of consensus OTUs onto a reference tree constructed from 16S rRNA sequences of type strains tagged as "good" quality, length between 1200bp and 1600bp in RDP 11.5 [69]. We performed multiple alignment of the sequences using the RDP's web-hosted alignment tool with default parameters [70]. To facilitate a good multiple alignment, we filtered out sequences with insertions seen in ≤ 3 other sequences. A reference tree was constructed using FastTree[71] version 2.1.7 SSE3 with the general-time-reversible maximum likelihood option. For phylogenetic placement, the aligned reference sequences were first trimmed to position 1045 to 1374 (corresponding to the region flanked by the 16S v4 primers) and a hidden Markov Model was learned using *hmmbuild* in HMMER v3.1[72]. OTU sequences were then aligned using *hmmalign* with the *-mapali* option. Finally, the aligned sequences were phylogenetically placed using *pplacer v1.1.alpha19* with default settings[73].

Alpha/Beta Ecological Diversity Analyses

Alpha diversity was assessed using normalized Shannon entropy. The Mann-Whitney U test was performed to assess significance of differences between healthy and dysbiotic microbiome cohorts, with correction for multiple hypotheses performed using the Benjamini-Hochberg procedure[74]. Beta diversity with the Bray-Curtis metric was used to assess the variation in the microbial compositions between healthy and dysbiotic donor microbiomes. Permutational Analysis of Variance (PERMANOVA) was performed to assess significance of differences between healthy and dysbiotic microbiome cohorts using *scikit-bio 0.5.6*.

Fold Change Analysis

Fold change analysis was performed using DESeq2 v1.3.2.0. All default options were used with features only kept if there were at least 100 reads (summing across all the samples used in the analysis) using the following commands

```
coldata$window<- factor(coldata$window)
dds <- DESeqDataSetFromMatrix(countData =
    cts,colData = coldata, design = ~window)
akeep <- rowSums(counts(dds)) >= 100
dds <- dds[akeep,]
dds<- DESeq(dds)
```

The scripts to perform this analysis are contained in https://github.com/gerberlab/MDSINE2_Paper/tree/master/deseq2. The fold changes were calculated in two ways, both using the default Wald test in the software. We first compared the steady state differences between the two cohorts by summing the reads over four consecutive measurements proceeding the first perturbation (steady state here is days $(16, 18, 21, 21.5)$) We also computed within cohort fold changes during the perturbations and compard them to the reads which were summed over four consecutive time points just before the perturbation began. In other words, we calculated the fold changes with respect to the "steady states" achieved just before the perturbation was applied. The HFD fold change was calculated by comparing days $(23, 23.5, 24, 25)$ to days $(16, 18, 21, 21.5)$, the vancomycin fold change was calculated by comparing days $(37, 37.5, 38, 39)$ to $(32, 33, 35, 35.5)$ and the gentamicin perturbation fold change was calculated by comparing days $(52,52.5,53,54)$ to $(46,47,50,50.5)$.

MDSINE2 analyses on H-cohort and D-cohort datasets

Model Inference

Day 0 and 0.5 samples were excluded from inferences, due to their very low overall bacterial concentrations. Models were inferred independently for the H-cohort and D-cohort datasets using 10,000 MCMC iterations after 5,000 burn-in steps. To assess convergence of MCMC chains, we used the $\hat{R}$ statistic[75] and confirmed values of $\hat{R}$<1.05 for the concentration parameter, the growth rates and the process variance, indicating sufficient mixing of MCMC chains. Inference took ~26.5 hours (6.3 sec/iteration) for the H-cohort and ~17.25 hours (4.1 sec/iteration) for the D-cohort on an Intel Xeon E5-2697V2 (2.70 GHz base frequency, 3.50 GHz max frequency) with an allocation of one core and 8gb of RAM separately for each task. Full details on inference are given in the Supplemental Information.

Benchmarking

We used implementations of the comparator methods provided in https://github.com/tyjo/clv. Following Joseph et al. (2020)[20], we trained the models using elastic-net regression, and in addition, we trained gLV using ridge regression to provide comparisons to earlier work [17,19]. Predictive performance of methods was assessed using a hold-one-subject-out cross validation procedure. Per fold, each method was provided data from all but one mouse in the cohort to infer model parameters. The inferred parameters were then used to forward simulate the trajectory of the held-out mouse, using the abundance at Day 1 as the initial condition. For comparator methods, the Runge-Kutta "rk45" procedure was used, as implemented in Joseph et al. (2020)[20]. For MDSINE2, each posterior sample was used to deterministically forward simulate (Equation 2 with no process variance), and the median of the distribution of simulations was used as the final forecast. The methods use different approaches to handle zeros in data. To make results as

comparable as possible, we used the following settings. For gLV-elastic net and gLV-ridge, we set the minimum value for taxa to 10^5 CFU/g, which is consistent with the limit of detection in our experiments. For gLV-ra and LRA, which are relative abundance methods, we set the minimum to 10^{-6} and for CLV, we set the additive offset $\epsilon = 10^{-6}$, consistent with the limit of detection for relative abundances in our experiments. These minimums were enforced both in data preprocessing and on the simulated trajectories, so that results remained comparable. For comparisons against methods that operate only on relative abundances (cLV, gLV-ra and LRA), we converted predictions of MDSINE2 or the gLV-based models to relative abundances. The following Root Mean Square Error metric was used:

$$\text{RMSLE}\left(X_{i,s}, \hat{X}_{i,s}\right) = \sqrt{\frac{1}{K}\sum_k \left(\log_{10} X_{i,s}(k) - \log_{10} \hat{X}_{i,s}(k)\right)^2},$$

where $X_{i,s}$ denotes the measurements for taxa *i* in the held-out mouse *s* and $\hat{X}_{i,s}$ are the respective forecast estimates. In order to compare the errors between MDSINE2 and other methods, we performed one-tailed Wilcoxon signed-rank testing. The paired data points used for the test are the RMSLEs associated with MDSINE2 and the RMSLEs associated with the comparator method for all the OTUs in the hold-out subjects.

Consensus Module Construction

Consensus modules were constructed by performing agglomerative clustering on the co-clustering probability matrix where the number of clusters is the median number of modules over the posterior. See our prior work for more details[29,30].

Taxonomic Enrichment Analysis

Using the consensus modules, we performed enrichment analysis at four taxonomic levels: family, order, class, and phylum. For details regarding taxonomic assignment, see section ASV aggregation into OTUs. The enrichment analysis was carried out using the hypergeometric test followed by Benjamini-Hochberg procedure for multiple hypothesis tests. The hypergeometric probability is defined as $\mathrm{P}(X = k) = \binom{M}{k}\binom{N-M}{n-k}/\binom{N}{n}$. N is the total number of OTUs used in the model, M is the total number of OTUs associated with a given taxonomy, n is the size of the interaction module, and k is the number of OTUs in the interaction module that is associated with the given taxonomy.

Phylogenetic Neighborhood Analysis

Phylogenetic Neighborhood Analysis (PNA) quantitatively assesses how phylogenetically similar but non-identical OTUs co-cluster. The input to PNA is a percent identity matrix $\boldsymbol{P}$ of pairwise OTU consensus sequence similarities and a matrix $\boldsymbol{C}$ of pairwise OTU co-clustering probabilities obtained from MCMC posterior inference. We compute $\boldsymbol{P}$ from the multiple alignment of the OTU consensus sequences and reference 16S v4 sequences as described above. PNA employs a coarsening procedure analogous to agglomerative clustering[76], but for PNA, agglomerates across both the phylogeny and co-clustering probabilities using average linkage measures. For two agglomerations of taxa, X and Y, the average phylogenetic distance is given by:

$$\boldsymbol{D}(X,Y) = \frac{1}{|X|\,|Y|}\sum_{x\in X, y\in Y}(1 - \boldsymbol{P}(x,y)) \qquad (i)$$

Similarly, the average of co-clustering probability is given by:

$$\overline{\boldsymbol{C}}(x,y) = \frac{1}{|X|\,|Y|}\sum_{x\in X, y\in Y}\boldsymbol{C}(x,y) \qquad (ii)$$

The high-level steps of the PNA algorithm are:

Initialize each OTU as its own agglomerate
Compute the distance $\boldsymbol{D}$ between the initial agglomerates using (i)
While the pairwise distance $\boldsymbol{D}$ between agglomerations is less than a threshold, *t*:
    Merge the agglomeration pairs with smallest distance (highest average percent identity)
    Update the pairwise distance $\boldsymbol{D}$ between the agglomerations using (i)
Return $\overline{\boldsymbol{C}}_{\text{Healthy}}$ and $\overline{\boldsymbol{C}}_{\text{D}}$, computed using (ii)

Here, $\overline{\boldsymbol{C}}$, are square matrices for each cohort, with the dimensions of each matrix equal to the number of agglomerations produced at that threshold, i.e., each row of the matrix denotes the probability of a given agglomeration unit co-clustering with other agglomeration units at the sequence distance threshold *t*. We used threshold values *t* in increments of 0.01 (1% identity) ranging from 0 to 0.26, where the final threshold corresponded to a single agglomeration containing all the OTUs. We report the correspondence between co-clustering probabilities of OTUs in the two cohorts, i.e., the degree to which phylogenetically similar OTUs co-cluster similarly across the healthy and dysbiotic donor microbiomes, as the mean of the Spearman correlations between the row vectors of each agglomeration unit. We generated a null distribution for co-clustering probabilities between the two conditions by randomly permuting the labels of $\boldsymbol{C}$ 10,000 times and performing PNA on each permuted matrix.

Simulation-based Stability Analysis

We simulated perturbations and assessed system recovery as follows:

1. Uniformly randomly select $O\alpha$ taxa to perturb, where *O* is the total number of taxa and α is a specified fraction to perturb.
2. For each selected taxon *i*, adjust its growth rate to $\boldsymbol{a}_{1,i} + \delta$ over a perturbation window of days 14 to 34.
3. Forward simulate taxa trajectories (as described above for Benchmarking) for 64 days, using the experimentally measured value at day 21 (e.g., a point at which consistent colonization is expected) as the initial condition.

This procedure was carried out over a range of parameters, $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$ and $\delta \in \{-0.5, -1.0, -1.5, -2.0\}$, with *N*=100 trials for each combination of parameters (α, δ). For each trial, 100 MCMC samples were used (every 100th sample out of the 10,000 samples) to perform forward simulations. For each trial *n*, the final state for taxa *i*, $\tilde{x}_i^{(n)}(\alpha,\delta)$, was computed as an average over 12 hours for the final simulated day. For comparison a baseline final state, $x_i^{(n)}$, in which the system was not perturbed, was computed in the same manner. Our measure of deviation from the baseline for each trial is given by:

$$d_{\text{ss}}\,(n;\alpha,\delta) = \text{mean}_i\left(\left|\log_{10}\left(\tilde{x}_i^{(n)}(\alpha,\delta) + \epsilon\right) - \log_{10}\left(x_i^{(n)} + \epsilon\right)\right|\right)$$

Here $\epsilon = 10^5$ is an additive constant (consistent with the limit of detection in our experiments).

Keystoneness

The keystoneness measure is computed by removing all the taxa for each module m, forward simulating trajectories (as described in

Benchmarking) for the remaining taxa over 100 days and comparing the final state of these trajectories to the final state with all taxa present in the ecosystem. As in our perturbation experiments for stability analysis, final states were computed as the mean of values over the last 12 hours in the final simulated day. To be precise, we obtain final state estimates $x^{(g)}$ (full system) and $\tilde{x}_m^{(g)}$ (system with module removed) for each MCMC step g, which are used to compute the keystoneness measure:

$$k(m) = -\,\text{mean}_g\left[\text{mean}_{i \notin m}\left(log_{10}\left(\tilde{x}_{mi}^{(g)} + \epsilon\right) - log_{10}\left(x_i^{(g)} + \epsilon\right)\right)\right],$$

where the subscript i denotes the taxon index. Just as in the simulation-based stability analysis, $\epsilon = 10^5$. Per this formulation, a positive k value indicates an overall decrease in the system on average (meaning m has a positive effect on other OTUs when present), while a negative k value indicates an increase (meaning m has a suppressive effect when present).

Contributions:

TEG: Gnotobiotic study design. Statistical model. Inference Algorithm. Software. Data analysis. Writing. Reviewing.
YK: Software. Data analysis. Writing methods.
SA: Software. Data analysis. Writing methods.
DEK: Software. Writing methods.
ND: Gnotobiotic study design and experiments.
RL: Gnotobiotic study design.
BB: Discussions with YK and critical review of the manuscript
JRA: Human donor sample collection and data acquisition. Critical review of the manuscript.
LB: Gnotobiotic study design and experiments. Detailed review of the manuscript.
GKG: Project conception. Gnotobiotic study design. Statistical model. Inference Algorithm. Software. Data analysis. Writing. Reviewing. Project management.

Declaration of financial interests:

No industry support was provided for this study.
TEG: None.
YK: None.
SA: None.
DK: None.
ND: None.
RL: None.
BB: None.
JRA consults for Finch Therapeutics, BMS, Pfizer, Janssen, Morphic, Iterative Scopes, Artugen, Servatus, Pandion, Merck, Baccain and has research support from Merck. No industry support was provided for this study.
LB is the inventor of patents for defined bacterial therapeutics for *C. difficile,* and is the SAB chair and a shareholder in ParetoBio, Inc. No industry support was provided for this study.
GKG is a shareholder in Kaleido Biosciences, Inc., and is on the SAB and is a shareholder in ParetoBio, Inc. His interests were reviewed and are managed by Brigham and Women's Hospital and Mass General Brigham in accordance with their conflict of interest policies.

Funding: Research reported in this publication was supported by DARPA BRICS HR0011-15-C-0094 (Gerber and Bry), NIH R01GM130777 (Gerber), NIH R35GM143056 (Gibson), NIH R21AI154075 (Gibson), NIH R35GM141861 (Berger), NIH P30DK056338 (Bry), NSF MTM2 2025512 (Gerber), the BWH President's Scholar Award (Gerber), and

Software and Data Availability: First time users wanting to explore the model and the data should start by visiting the main repo for this paper https://github.com/gerberlab/MDSINE2_Paper. All sequencing data for this study are available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA784519. Post processed reads, qPCR measurements values, and sample collection times are all provided in https://github.com/gerberlab/MDSINE2_Paper/tree/master/datasets/gibson. The core MDSINE2 package is available at https://github.com/gerberlab/MDSINE2. Pickled files containing the posterior samples from running the model on our data are available at https://doi.org/10.5281/zenodo.5527256.
Scripts showing how to use MDSINE2 and reproduce the main text figures can be found in https://github.com/gerberlab/MDSINE2_Paper/tree/master/google_colab.

References

1 Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nature Medicine* **24**, 392-400, doi:10.1038/nm.4517 (2018).
2 Sampson, T. R. *et al.* Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's disease. *Cell* **167**, 1469-1480. e1412 (2016).
3 Yano, J. M. *et al.* Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* **161**, 264-276 (2015).
4 Li, B., Selmi, C., Tang, R., Gershwin, M. E. & Ma, X. The microbiome and autoimmunity: a paradigm from the gut–liver axis. *Cellular & Molecular Immunology* **15**, 595-609, doi:10.1038/cmi.2018.7 (2018).
5 Pascale, A. *et al.* Microbiota and metabolic diseases. *Endocrine* **61**, 357-371, doi:10.1007/s12020-018-1605-5 (2018).
6 Shreiner, A. B., Kao, J. Y. & Young, V. B. The gut microbiome in health and in disease. *Curr Opin Gastroenterol* **31**, 69 (2015).
7 Gerber, G. K. The dynamic microbiome. *FEBS letters* **588**, 4131-4139 (2014).
8 Gonze, D., Coyte, K. Z., Lahti, L. & Faust, K. Microbial communities as dynamical systems. *Current opinion in microbiology* **44**, 41-49 (2018).
9 May, R. M. Will a large complex system be stable? *Nature* **238**, 413-414 (1972).
10 Goh, B. S. Global Stability in Many-Species Systems. *The American Naturalist* **111**, 135-143 (1977).
11 Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, competition, and stability. *Science* **350**, 663-666 (2015).
12 Rao, C. *et al.* Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature* **591**, 633-638, doi:10.1038/s41586-021-03241-8 (2021).
13 Lee, T. I. *et al.* Transcriptional regulatory networks in Saccharomyces cerevisiae. *science* **298**, 799-804 (2002).
14 Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827 (2002).
15 Barabasi, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature reviews genetics* **5**, 101-113 (2004).
16 Franzosa, E. A. *et al.* Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology* **4**, 293-305, doi:10.1038/s41564-018-0306-4 (2019).
17 Stein, R. R. *et al.* Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. *PLoS Comput Biol* **9** (2013).
18 Fisher, C. K. & Mehta, P. Identifying Keystone Species in the Human Gut Microbiome from Metagenomic Timeseries Using Sparse Linear Regression. *PLoS ONE* **9**, e102451 (2014).

19 Bucci, V. *et al.* MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. *Genome Biol* **17**, doi:10.1186/s13059-016-0980-6 (2016).

20 Joseph, T. A., Shenhav, L., Xavier, J. B., Halperin, E. & Pe'er, I. Compositional Lotka-Volterra describes microbial dynamics in the simplex. *PLOS Computational Biology* **16**, e1007917, doi:10.1371/journal.pcbi.1007917 (2020).

21 Thanissery, R., Winston, J. A. & Theriot, C. M. Inhibition of spore germination, growth, and toxin activity of clinically relevant C. difficile strains by gut microbiota derived secondary bile acids. *Anaerobe* **45**, 86-100 (2017).

22 Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *The ISME Journal* **6**, 1535-1543, doi:10.1038/ismej.2012.4 (2012).

23 Cinquin, C., Le Blay, G., Fliss, I. & Lacroix, C. New three-stage in vitro model for infant colonic fermentation with immobilized fecal microbiota. *FEMS Microbiology Ecology* **57**, 324-336, doi:10.1111/j.1574-6941.2006.00117.x (2006).

24 Bakshi, S. *et al.* Tracking bacterial lineages in complex and dynamic environments with applications for growth control and persistence. *Nature Microbiology* **6**, 783-791, doi:10.1038/s41564-021-00900-4 (2021).

25 Jalili-Firoozinezhad, S. *et al.* A complex human gut microbiome cultured in an anaerobic intestine-on-a-chip. *Nature Biomedical Engineering* **3**, 520-531, doi:10.1038/s41551-019-0397-0 (2019).

26 Lavin, R., DiBenedetto, N., Yeliseyev, V., Delaney, M. & Bry, L. Gnotobiotic and Conventional Mouse Systems to Support Microbiota Based Studies. *Curr Protoc Immunol* **121**, e48, doi:10.1002/cpim.48 (2018).

27 Faith, J. J., Ahern, P. P., Ridaura, V. K., Cheng, J. & Gordon, J. I. Identifying gut microbe–host phenotype relationships using combinatorial communities in gnotobiotic mice. *Science translational medicine* **6**, 220ra211-220ra211 (2014).

28 Planer, J. D. *et al.* Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature* **534**, 263-266 (2016).

29 Gerber, G. K., Onderdonk, A. B. & Bry, L. Inferring dynamic signatures of microbes in complex host ecosystems. *PLoS Comput Biol* **8**, e1002624, doi:10.1371/journal.pcbi.1002624 (2012).

30 Creswell, R. *et al.* High-resolution temporal profiling of the human gut microbiome reveals consistent and cascading alterations in response to dietary glycans. *Genome Medicine* **12**, 59, doi:10.1186/s13073-020-00758-x (2020).

31 Simberloff, D. & Dayan, T. The Guild Concept and the Structure of Ecological Communities. *Annual Review of Ecology and Systematics* **22**, 115-143, doi:10.1146/annurev.es.22.110191.000555 (1991).

32 Hofman, J. M. & Wiggins, C. H. Bayesian approach to network modularity. *Physical review letters* **100**, 258701 (2008).

33 Segal, E., Friedman, N., Kaminski, N., Regev, A. & Koller, D. From signatures to models: understanding cancer using microarrays. *Nature genetics* **37**, S38-S45 (2005).

34 Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583, doi:10.1038/nmeth.3869 (2016).

35 Cao, H. T., Gibson, T. E., Bashan, A. & Liu, Y. Y. Inferring human microbial dynamics from temporal metagenomics data: Pitfalls and lessons. *Bioessays* **39**, doi:10.1002/bies.201600188 (2017).

36 Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature* **517**, 205-208, doi:10.1038/nature13828 (2015).

37 Girinathan, B. P. *et al.* In vivo commensal control of Clostridioides difficile virulence. *Cell Host & Microbe* (2021).

38 Kass, R. E. & Raftery, A. E. Bayes Factors. *Journal of the American Statistical Association* **90**, 773-795, doi:10.1080/01621459.1995.10476572 (1995).

39 Carmody, R. N. *et al.* Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* **17**, 72-84, doi:10.1016/j.chom.2014.11.010 (2015).

40 Gibson, T. E., Bashan, A., Cao, H.-T., Weiss, S. T. & Liu, Y.-Y. On the Origins and Control of Community Types in the Human Microbiome. *PLOS Computational Biology* **12**, e1004688, doi:10.1371/journal.pcbi.1004688 (2016).

41 Doyle, J. C., Francis, B. A. & Tannenbaum, A. R. *Feedback control theory*. (Courier Corporation, 2013).

42 Gibson, T. E. Sign Stability via Root Locus Analysis. *arXiv preprint arXiv:1512.06026* (2015).

43 Hughes, R. L. *et al.* Resistant Starch Type 2 from Wheat Reduces Postprandial Glycemic Response with Concurrent Alterations in Gut Microbiota Composition. *Nutrients* **13**, 645 (2021).

44 Forbes, J. D. *et al.* A comparative study of the gut microbiota in immune-mediated inflammatory diseases—does a common dysbiosis exist? *Microbiome* **6**, 221, doi:10.1186/s40168-018-0603-4 (2018).

45 Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655-662, doi:10.1038/s41586-019-1237-9 (2019).

46 Kaakoush, N. O. Sutterella Species, IgA-degrading Bacteria in Ulcerative Colitis. *Trends in Microbiology* **28**, 519-522, doi:10.1016/j.tim.2020.02.018 (2020).

47 Schirmer, M. *et al.* Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. *Cell host & microbe* **24**, 600-610. e604 (2018).

48 Paramsothy, S. *et al.* Multidonor intensive faecal microbiota transplantation for active ulcerative colitis: a randomised placebo-controlled trial. *The Lancet* **389**, 1218-1228 (2017).

49 Hyams, J. S. *et al.* Clinical and biological predictors of response to standardised paediatric colitis therapy (PROTECT): a multicentre inception cohort study. *The Lancet* **393**, 1708-1720 (2019).

50 Bar-Joseph, Z. *et al.* Computational discovery of gene modules and regulatory networks. *Nature Biotechnology* **21**, 1337-1342, doi:10.1038/nbt890 (2003).

51 Gerber, G. K. *Computational discovery of gene modules, regulatory networks and expression programs* PhD thesis, Harvard and MIT Division of Health Sciences and Technology, (2007).

52 Gerber, G. K., Dowell, R. D., Jaakkola, T. S. & Gifford, D. K. Automated discovery of functional generality of human gene expression programs. *PLoS Comput Biol* **3**, e148, doi:10.1371/journal.pcbi.0030148 (2007).

53 Baxter, N. T. *et al.* Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary Interventions with Three Fermentable Fibers. *mBio* **10**, e02566-02518, doi:doi:10.1128/mBio.02566-18 (2019).

54 Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446-450 (2013).

55 Ziesack, M. *et al.* Engineered interspecies amino acid cross-feeding increases population evenness in a synthetic bacterial consortium. *Msystems* **4**, e00352-00319.

56 Neal, R. M. Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics* **9**, 249-265 (2000).

57 Antoniak, C. E. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The annals of statistics*, 1152-1174 (1974).

58 Escobar, M. D. & West, M. Bayesian density estimation and inference using mixtures. *Journal of the american statistical association* **90**, 577-588 (1995).

59 McMurdie, P. J. & Holmes, S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS computational biology* **10**, e1003531 (2014).

60 Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357-362 (2020).

61 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods* **17**, 261-272 (2020).
62 Lam, S. K., Pitrou, A. & Seibert, S. in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC.* 1-6.
63 Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in science & engineering* **9**, 90-95 (2007).
64 Waskom, M. & team, S. d. mwaskom/seaborn. (2020). <https://doi.org/10.5281/zenodo.592845>.
65 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504 (2003).
66 Hsu, B. B. *et al.* Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host Microbe* **25**, 803-814 e805, doi:10.1016/j.chom.2019.05.001 (2019).
67 Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology* **79**, 5112-5120 (2013).
68 Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**, 5261-5267, doi:10.1128/aem.00062-07 (2007).
69 Cole, J. R. *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic acids research* **37**, D141-D145 (2009).
70 Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335-1337 (2009).
71 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* **5**, e9490 (2010).
72 Eddy, S. R. Accelerated profile HMM searches. *PLoS computational biology* **7**, e1002195 (2011).
73 Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC bioinformatics* **11**, 1-16 (2010).
74 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300 (1995).
75 Gelman, A. *et al. Bayesian data analysis*. (Chapman and Hall/CRC, 2013).
76 Leskovec, J., Rajaraman, A. & Ullman, J. D. *Mining of massive data sets*. (Cambridge university press, 2020).