Microbial dynamics inference at ecosystem-scale

Travis E. Gibson[1,2,4,7,#], Younhun Kim[1,2,6], Sawal Acharya[1], David E. Kaplan[1], Nicholas DiBenedetto[3], Richard Lavin[3], Bonnie Berger[2,5,6,7], Jessica R. Allegretti[4,8], Lynn Bry[3,4], and Georg K. Gerber[1,3,4,5,#]

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

Abstract: Dynamical systems models are a powerful tool for analyzing interactions, stability, resilience, and other key properties in biomedically important microbial ecosystems, such as the gut microbiome. Challenges to modeling and inference in this setting include the large number of species present, and data sparsity/noise characteristics. Here, we introduce a Bayesian statistical method, the Microbial Dynamical Systems Inference Engine 2 (MDSINE2), which infers compact and interpretable ecosystems-scale dynamical systems models from microbiome time-series data. We model microbial dynamics as stochastic processes driven by inferred interaction modules, or groups of microbes with similar interaction structure and responses to perturbations. Additionally, we model the noise characteristics of sequencing and qPCR measurements to provide uncertainty quantification for all outputs. To evaluate MDSINE2, and provide a benchmarking resource for the community, we generated the most densely sampled microbiome time-series to date, which consists of a cohort of mice that received fecal transplants from a human donor and were then subjected to dietary and antibiotic perturbations. Benchmarking on simulated and real data demonstrate that MDSINE2 significantly outperforms state-of-the-art methods, and moreover identifies interaction modules that shed new light on ecosystems-scale interactions in the gut microbiome. We provide MDSINE2 as an open-source Python package at: https://github.com/gerberlab/MDSINE2.

Introduction

Microbiomes are inherently dynamic[2], changing over time due to both microbial interactions, as well as responses to external perturbations. Dynamics of a microbiome reveals important information not only about the individual microbial constituents, but also about how the ecosystem as a whole behaves; for instance, unstable responses of the ecosystem to perturbations can be indicative of an inability to maintain homeostatic function[3]. Mathematical models of dynamical systems have a long history in ecology and biomedicine, and have led to many insights, including for microbial ecosystems[4]. 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[5-8]; topological properties of the interaction network such as motifs[9-11]; and *in silico* forecasts of the system such as "knock-outs" of taxa or responses to perturbations not yet experimentally studied**.**

However, the scale and complexity of microbiomes, as well as limitations of measurement modalities, present challenges to modeling and inferring dynamical systems models from data. The gut microbiome contains hundreds of ecologically diverse yet interacting microbes. Indeed, there is increasing recognition that this interaction structure is critically important, driving behaviors such as whether so-called pathobiont bacteria will cause disease or remain harmless in the host[12]. A well-established modeling framework, which we and others have employed, uses the generalized Lotka-Volterra (gLV) equations[1,13-15] 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[1], 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. This large number of interaction parameters must 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[16,17], but are orders of magnitude lower than the predominant taxa in the gut. Sufficiently rich experimental data is also critical for model inference[18]. Perturbations are essential to analyze how components of the system interact and to assess stability and other key properties; data at equilibrium in a single or small number of conditions cannot be used to infer these properties of the system.

Here, we provide two new resources to the community to address the challenge of inferring microbial dynamics at ecosystem-scale: (1) a fully Bayesian computational method, the Microbial Dynamical Systems Inference Engine 2 (MDSINE2), implemented as an open-source software package, and (2) a densely sampled gut microbiome time-series dataset for benchmarking and other analyses. MDSINE2 has several advantages over state-of-the-art methods, including our own prior contributions[1,13-15]. Our method is the first to both explicitly model measurement noise and fully propagate this information during inference, which has distinct advantages for quantifying uncertainty of predictions and more accurately capturing behavior of low abundance species in the ecosystem. Further, MDSINE2 automatically learns interaction modules, or groups of bacteria with similar interaction structure and responses to perturbations, which allows our model to scale to large ecosystems while maintaining interpretability. We make MDSINE2 available to the community as a fully open-source Python package, and also provide extensive Colab tutorials. To provide a community resource for benchmarking MDSINE2 and other methods, we generated, to our knowledge, the most temporally densely sampled gut microbiome dataset to date that includes perturbations essential for dynamical systems inference. For this dataset, we developed a cohort of "humanized" gnotobiotic mice via fecal transplantation from a healthy human donor and then subjected mice to series of dietary and antibiotic perturbations, while collecting an average of 77 fecal samples per mouse over 65 days. Samples were then analyzed with both high-throughput 16S rRNA gene amplicon sequencing to estimate microbial relative abundances and 16S rRNA gene qPCR to quantitate total microbial concentrations, with physical replicates included to calibrate computational noise models.

Results

MDSINE2 is an open-source computational tool for inferring dynamical systems models of microbiomes at scale

To infer accurate and interpretable large-scale dynamical systems models from microbiome time-series data, we developed MDSINE2 (Figure 1), a fully Bayesian machine learning model. Inference is performed using a custom Markov Chain Monte Carlo (MCMC) algorithm to approximate the posterior probability, which is implemented in a software package that leverages vectorization and just-in-time compilation technologies for high efficiency. Inputs to the MDSINE2 software are: (1) time-series measurements of bacterial abundances in the form of counts (e.g., 16S rRNA gene amplicon or shotgun metagenomics data), (2) total bacterial concentrations (e.g., from 16S rRNA gene qPCR measurements),

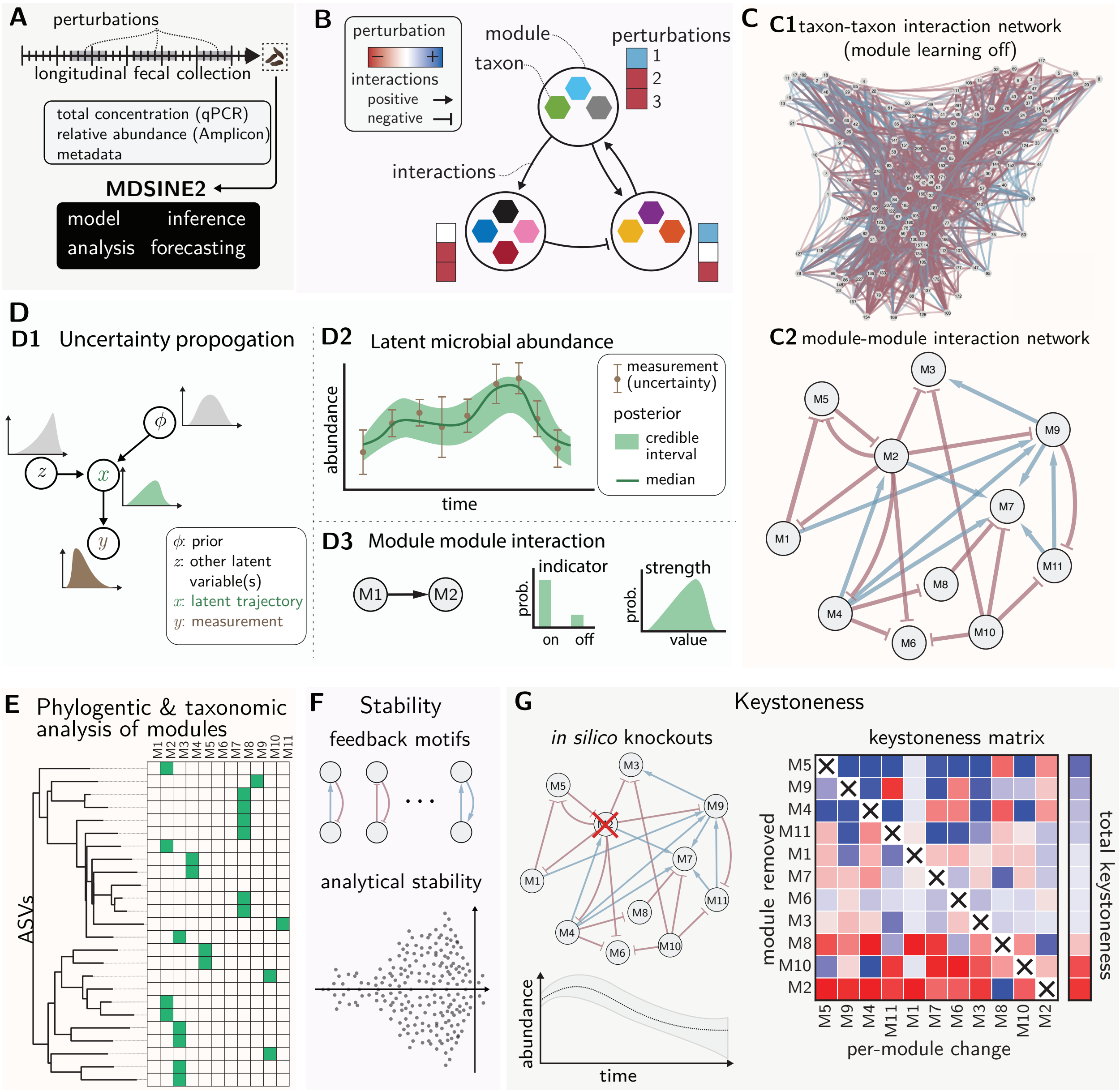

A
perturbations
longitudinal fecal collection
total concentration (qPCR)
relative abundance (Amplicon)
metadata
MDSINE2
model
inference
analysis
forecasting
B
perturbation
−
+
interactions
positive
negative
module
taxon
perturbations
1
2
3
interactions
C
C1 taxon-taxon interaction network
(module learning off)
C2 module-module interaction network
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
D
D1 Uncertainty propogation
ϕ
z
x
y
ϕ: prior
z: other latent variable(s)
x: latent trajectory
y: measurement
D2 Latent microbial abundance
abundance
time
measurement (uncertainty)
posterior
credible interval
median
D3 Module module interaction
M1
M2
indicator
prob.
on
off
strength
prob.
value
E
Phylogentic & taxonomic analysis of modules
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
ASVs
F
Stability
feedback motifs
analytical stability
G
Keystoneness
in silico knockouts
M5
M3
M9
M7
M1
M11
M8
M4
M6
M10
abundance
time
keystoneness matrix
module removed
M5
M9
M4
M11
M1
M7
M6
M3
M8
M10
M2
per-module change
total keystoneness

Figure 1. Schematic of the MDSINE2 method for inferring interpretable dynamical systems models of microbiomes at scale. (A) Input data are measurements of total bacterial concentration (e.g., 16S rRNA gene qPCR) and measurements of taxa abundances (e.g., 16S rRNA gene amplicon sequencing). Measurements are obtained from studies in which the microbiome undergoes perturbations, providing sufficiently rich information for inference. **(B)** MDSINE2 infers dynamical systems models from data with the option of automatically learning interaction modules, or groups of taxa that share the same interactions with other modules and perturbations. This is a more compact representation that is more readily interpretable than learning interactions among all microbes. **(C)** Example microbial interaction networks for the same number of taxa without module learning (C1) and with module learning (C2)**. (D)** MDSINE is fully Bayesian and propagates error throughout the model (D1), providing estimates of uncertainty for all variables, (e.g. (D2) latent trajectory along with measurements and their uncertainty, (D3) indicator and interaction strength for ecological interactions. The software provides a variety of tools for analysis and visualization of the inferred dynamical system, including: **(E)** analyses of taxonomic composition and phylogeny of modules, **(F)** formal analyses of ecosystem stability and interaction motifs, and **(G)** keystoneness (quantitative impact on the ecosystem when modules are removed).

and (3) associated metadata for the samples (Figure 1A). The software also provides a variety of tools for interrogating 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 (Figure 1E,F,G). The core MDSINE2 software is maintained in the Gerber Lab repo https://github.com/gerberlab/MDSINE2 while the code to reproduce the figures for this paper along with Google Colab notebooks for exploring the model are maintained at https://github.com/gerberlab/MDSINE2_Paper.

Our computational model is based on generalized Lotka-Volterra dynamics, but with several key innovations compared to prior state of the art[1,13-15]. First, MDSINE2 explicitly models the measurement uncertainty associated with microbiome sequencing and qPCR measurements. Unlike our own previous method[1], which handled data denoising and dynamical systems inference in two separate steps, MDSINE2 fully propagates uncertainty through the model (Figure 1D) and provides quantitative measures (e.g., Bayes factors) of uncertainty for all model parameters. Second, MDSINE2 includes stochastic effects in dynamics, which capture non-deterministic fluctuations in microbial trajectories that occur due to unmeasured effects on the ecosystem. Third, MDSINE2 extends the gLV model to include automatically learn *interaction modules* (Figure 1B), which we define as groups of taxa that share common interaction structure (i.e., are 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[19,20] and theoretical ecology concepts such as guilds, or groups of taxa that utilize resources in a similar way[21]. Modular structure (Figures 1B and 1C2) reduces the complexity of the system to be analyzed, which has the potential to increase human interpretability[22,23]. Further, scalability is enabled: 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 in the gLV equations) to order quadratic in the number of modules (which scales logarithmically with the number of taxa)[24]. The number of modules is treated probabilistically with full uncertainty quantification, using a nonparametric Bayesian framework, and learned from the data, alleviating the need for the user to pre-specify this information. See Methods §1 and Supplemental Text §1 for details on the model and Methods §4.1 and Supplemental Text §2 for details on inference.

MDSINE2 outperforms state-of-the-art models on existing in silico benchmark datasets

To provide an initial point of comparison, we assessed MDSINE2 against our prior method, as well as two state-of-the-art gLV-based methods: ridge regression (gLV-L2) and elastic net regression (gLV-net)[13], using an *in silico* benchmarking standard we previously published[1] (Figure 2). Briefly, the dataset simulates gLV dynamics for a small ecosystem of 10 taxa with parameters resembling those learned on real experimental data (Figure 2A and Methods §4.3). We note that we did not compare against methods that use relative abundance as input, because these methods cannot estimate ground truth growth rates or

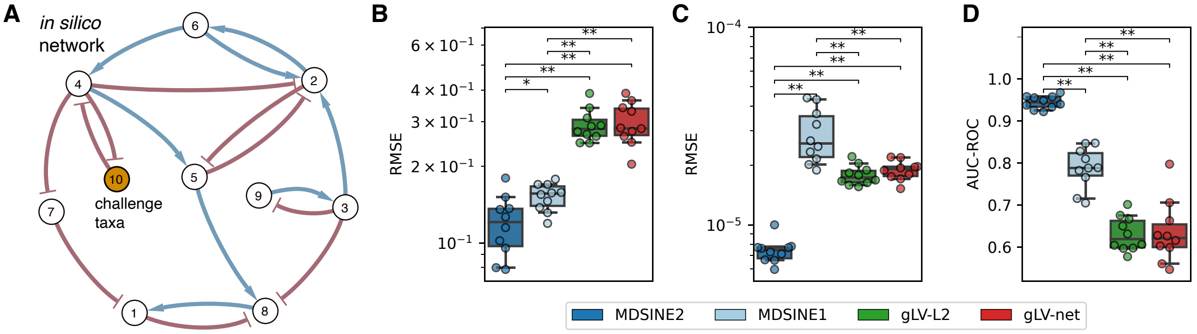


Figure 2: MDSINE2 outperforms state-of-the-art methods on the a synthetic data benchmarking standard. (A) Underlying dynamical systems network topology[1], which was used to simulate data under a gLV model. Root mean-squared error (RMSE) for **(B)** growth rates (lower is better) **(C)** interaction strengths (lower is better), and **(D)** Area under the receiver operating curve (AUC-ROC) for network topology (higher is better, 1 is maximum). * p<0.05, ** p<0.01.

interaction strength, requiring one to provide a rescaling factor (which is unidentifiable when trained on relative abundance data). MDSINE2 significantly outperformed the comparator methods, including our own prior method, on all metrics: inference of growth rates, interaction strengths, and the ground-truth network structure (Figure 2B,C,D and Supplemental Table 1). These results suggest that the new modeling capabilities in MDSINE2 provide significant performance boosts on critical dynamical systems inference tasks. However, this *in silico* dataset simulates only a small ecosystem, and moreover is not expected to fully recapitulate characteristics of real data. Thus, we next sought to generate a new experimental dataset for fully benchmarking the methods and to provide insights into the dynamics of real, large-scale microbial ecosystems.

High-temporal resolution gut microbiome study with serial perturbations is a resource for microbial dynamical systems inference benchmarking and analysis

Given the importance of sufficient temporal resolution and perturbation information for inferring dynamical systems models from data[18], and the dearth of appropriate datasets available to researchers, we created a new dataset to serve as a benchmarking and analysis resource for the community. To generate the dataset, we created a cohort of “humanized” mice (Figure 3A) by performing a fecal microbiota transplantation from a healthy human donor into germ-free mice (n=4). Gnotobiotic mice, in which bacteria-free mice are inoculated with micro-organisms, are a compelling experimental system that

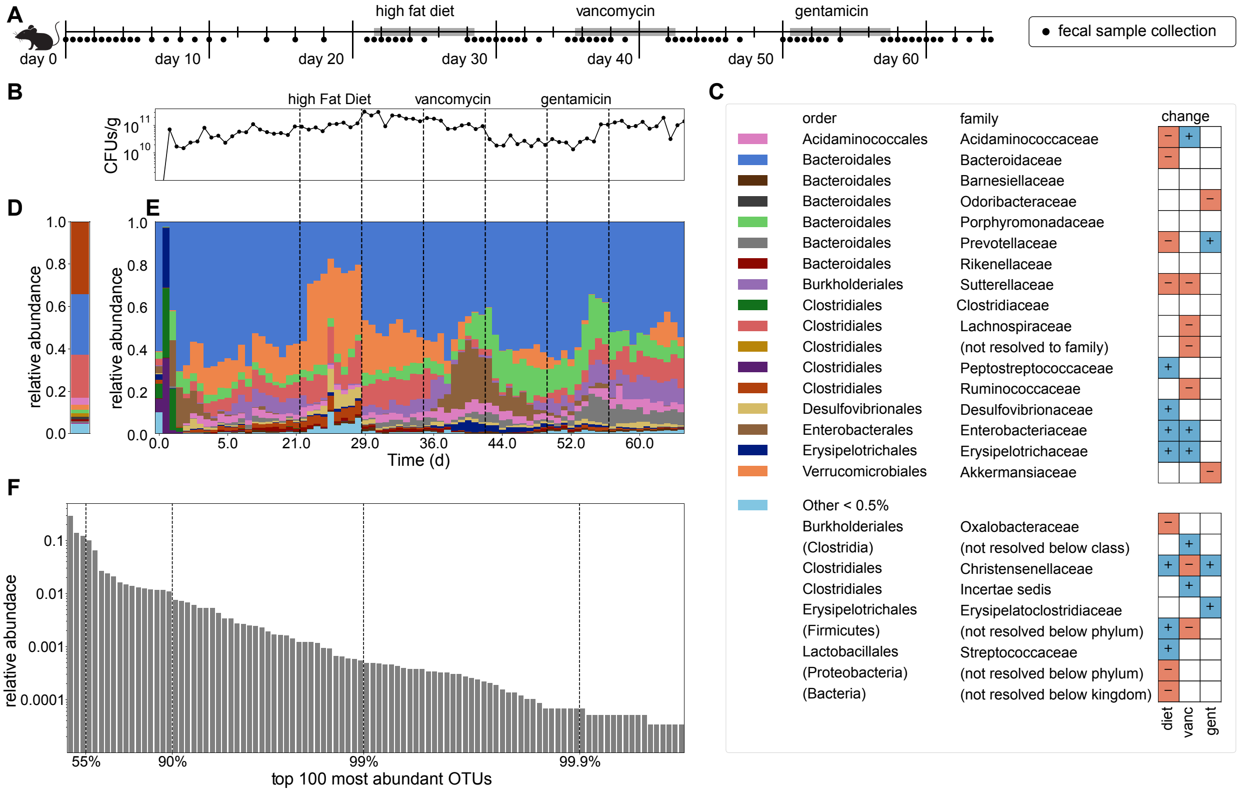


Figure 3: High-temporal resolution gnotobiotic mice colonization and perturbation studies using a human donor microbiome shows reproducible differential responses to perturbations. (A) Experimental design for the study (n = 4 mice) with an average of 77 serial fecal samples/mouse. **(B)** Average total bacterial concentrations in serial fecal samples. **(C)** Legend and taxonomy for panels D and E along with significant differential abundances for each taxonomic group across the three perturbations (see Supplemental Table 2 for *p*-values). **(D)** Relative abundance of microbes in human donor sample. **(E)** Relative abundances of microbes in serial fecal samples, averaged over the biological replicates. **(F)** Relative abundances of the top 100 most abundant ASVs across all mouse samples.

has been shown to sustain the vast majority of bacterial species found in native human microbiomes[25] and recapitulate important aspects of human physiology and pathology[1,26,27]. From the standpoint of dynamical systems inference, gnotobiotic models allow for targeted strong perturbations and high-temporal resolution fecal sampling.

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 versus complex carbohydrate utilizers and bacteria susceptible or resistant to different antibiotics). Mice were separately housed and fecal samples were collected over a 65-day duration, with an average of 77 samples per mouse (Figure 3A). Samples were interrogated for relative abundance via 16S rRNA amplicon (Figure 3D, E) sequencing and total bacterial concentration via qPCR using a universal 16S rDNA primer (Figure 3B).

The resulting 24,027,980 million sequencing reads were bioinformatically processed using DADA2[28], resulting in 1088 Amplicon Sequence Variants (ASVs). To yield a dataset with high-quality time-series information for dynamical systems inference tasks[18], taxa that were not present at ≥0.01% relative abundance for seven consecutive time-points in at least two mice were filtered out, resulting in 141 ASVs for inference benchmarking; of note, this is more than 10X the scale of the original benchmarking dataset we provided to the community[1]. Figure 3F displays the relative abundances of the top 100 most prevalent taxa, illustrating the commonly seen phenomenon in gut microbiome datasets of orders-of-magnitude ranges in abundance among taxa (i.e., the top three most abundant taxa account for 55% of reads, and 90% of the reads account for only 17 taxa).

As expected from the experimental design, phylogenetically diverse groups of taxa underwent reproducible differential shifts from the applied perturbations (Figure 3C,E Supplemental Figure 1, Supplemental Tables 2,3). Standard univariate fold-change analyses[29] demonstrated significant shifts across mice at the Phylum level, including decreases of Bacteroidetes and increases of Firmicutes on the high fat diet; decreases of Firmicutes and Actinobacteria, and increases of Proteobacteria on vancomycin; and increases in Firmicutes and Actinobacteria on gentamicin. At the Family level, significant shifts across mice were also seen in many taxa, including decreases in Prevotellaceae and Sutterellaceae, and increases in Peptostreptococcaceae, Desulfovibrionaceae, and Erysipelotrichaceae on the high fat diet; decreases in Sutterellaceae, Lachnospiraceae, and Ruminococcaceae and increases in Enterobacteriaceae and Erysipelotrichaceae on vancomycin; and decreases in Odoribacteraceae and Akkermansia and increases in Prevotellaceae on gentamicin. These results, showing consistent colonization with 141 ASVs in mice and reproducible differential responses to perturbations, coupled with the temporal resolution of the data, suggest its utility for inferring dynamical systems models, which we demonstrate in the subsequent sections.

MDSINE2 outperformed state-of-the-art methods in forecasting microbial dynamics at scale

Using the experimental dataset we generated, we evaluated MDSINE2 and the gLV ridge and elastic net comparator methods for their ability to forecast held-out microbial concentrations. We additionally evaluated MDSINE2 without interaction modules (MDSINE2^{-M}), to assess the impact of this feature on model performance and also to more directly compare our method to the state-of-the-art methods, which do not infer modules. 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 (RMSE) over the time-series, e.g., a measure of the difference between the predicted and ground-truth measurement.

Our method significantly outperformed both comparator methods on this benchmark for training with total bacterial concentration (Figure 4A, Supplemental Table 4). MDSINE2^{-M}, which does not learn

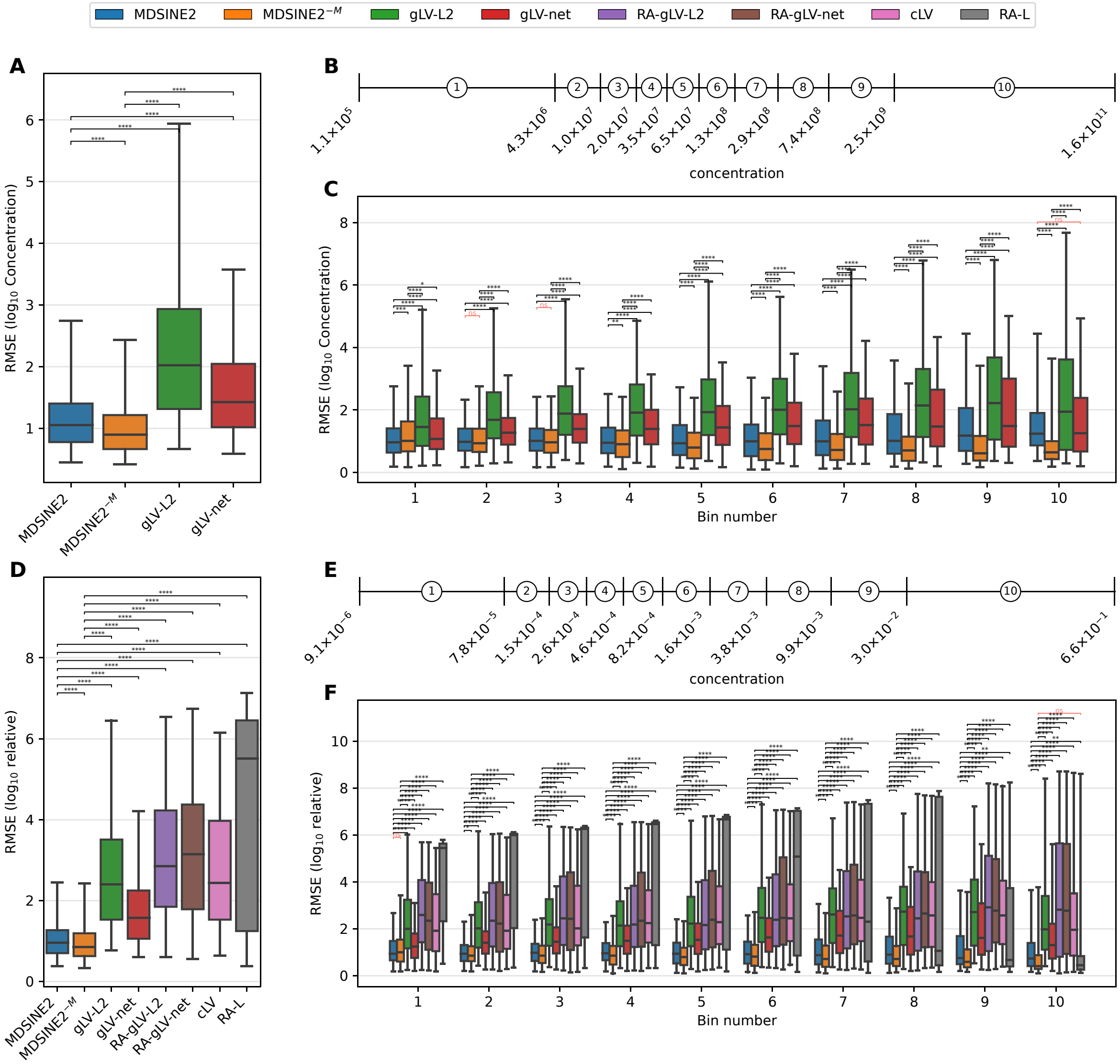


Figure 4: MDSINE2 outperforms state-of-the-art methods in forecasting microbial concentrations and relative abundances. Results are for cross-fold validation of forecasts. Models were trained on three of four mice, then given the initial condition of the hold out mouse time-series, and then the inferred model was simulated forwards and the output was compared to the entire held out time-series. All tests are two-sided Wilcoxon signed-rank tests. Non-significant tests are denoted with red bracket. Errors are calculated from forecasted abundances with respect to hold out data across all folds and for all taxa with non-zero reads within each sample. **(A-C)** Results for models trained on microbial concentrations. **(A)** Forecasting error for all non-zero taxa samples. **(B)** Concentration ranges for deciles. **(C)** Forecasting error grouped by decile of holdout sample taxa concentrations. **(D-F)** Results comparing models trained on concentrations and relative abundances. **(D)** Forecasting error for all non-zero taxa samples. **(E)** Relative abundance ranges for deciles. **(F)** Forecasting error grouped by decile of holdout sample taxa relative abundance. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$, ns: not significant ($p>0.05$).

modules and instead encodes parameters for each pairwise interaction among taxa (19,740 parameters from 141 ASVs), showed slight but statistically significant better forecasting accuracy over MDSINE2 (272 parameters from a median of 17 interaction modules). This result is consistent with our prior finding that model constraints can impact forecasting performance to some extent[1]. However, the actual gap in performance between MDSINE2^{-M} and MDSINE was quite minor, suggesting that the much more compact dynamical system representation learned by MDSINE2 still captures the system behavior quite accurately.

Given the large range in taxa abundances, we were interested in assessing how predictive performance varied with abundance, and in particular the impact on lower abundance taxa, which have an inherently larger relative variance[30] due to being measured with fewer sequencing reads. MDSINE2^{-M} significantly outperformed the two comparator methods at all abundances (Figure 4B,C and Supplemental Table 5), including down to 10^5 CFU/g, which is near the limit of detection for the dataset. MDSINE2 similarly significantly outperformed the comparator methods in all but the most abundant bin (concentrations greater than $2.5\text{x}10^9$ CFU/g). Interestingly, while MDSINE2^{-M} showed minor but statistically significant outperformance of MDSINE2 in the top 7 deciles (concentrations greater than $2\text{x}10^7$ CFU/g) there was no significant performance difference in the second and third lowest deciles ranging from $4.3\text{x}10^6$ CFU/g to $2\text{x}10^7$ CFU/g, and MDSINE2 significantly outperformed MDSINE2^{-M} for the lowest decile $1.1\text{x}10^5$ CFU/g to $4.3\text{x}10^6$ CFU/g. These results suggest that interaction modules capture dynamics well across the range of taxa abundances and may even improve accuracy for the least abundant taxa. We note that none of the methods, including our own, can predict exactly zero abundances. Analysis of cases in which the held-out sample ASVs have zero reads show poor predictive performance for all methods, which all generally predict values near the limit of detection of 10^5-10^6 CFU/g, as expected (Supplemental Figure 2A). Interestingly the method that had the best performance for the zero read cases, gLV-L2, had the worst performance in all other cases.

To assess the role that bacterial concentration information gives for model forecasting, we compared the performance of the four models just described against other state-of-the-art methods that learn from relative abundance information alone. These methods include: compositional Lotka Volterra (cLV)[15], linear dynamics trained on relative abundances (RA-L), gLV ridge trained on relative abundances (RA-gLV-L2), and gLV elastic net trained on relative abundances (RA-gLV-net). Overall, MDSINE2 and MDSINE2^{-M} outperformed the comparator methods, and MDSINE2^{-M} ultimately had the best performance (Figure 4D,E,F and Supplemental Tables 6, 7). However, we found that methods trained on data that included the qPCR measurements significantly outperformed methods trained on only relative abundance information, with the only exception being gLV-L2 compared to cLV (Figure 4F, Supplemental Table 6). These results highlight a well-studied issue with relative abundance data, that increases in one taxa's abundance are indistinguishable from other taxa's abundances simultaneously decreasing[31], which has implications for identifiability (the ability to learn model parameters from data) from a dynamical systems perspective[18].

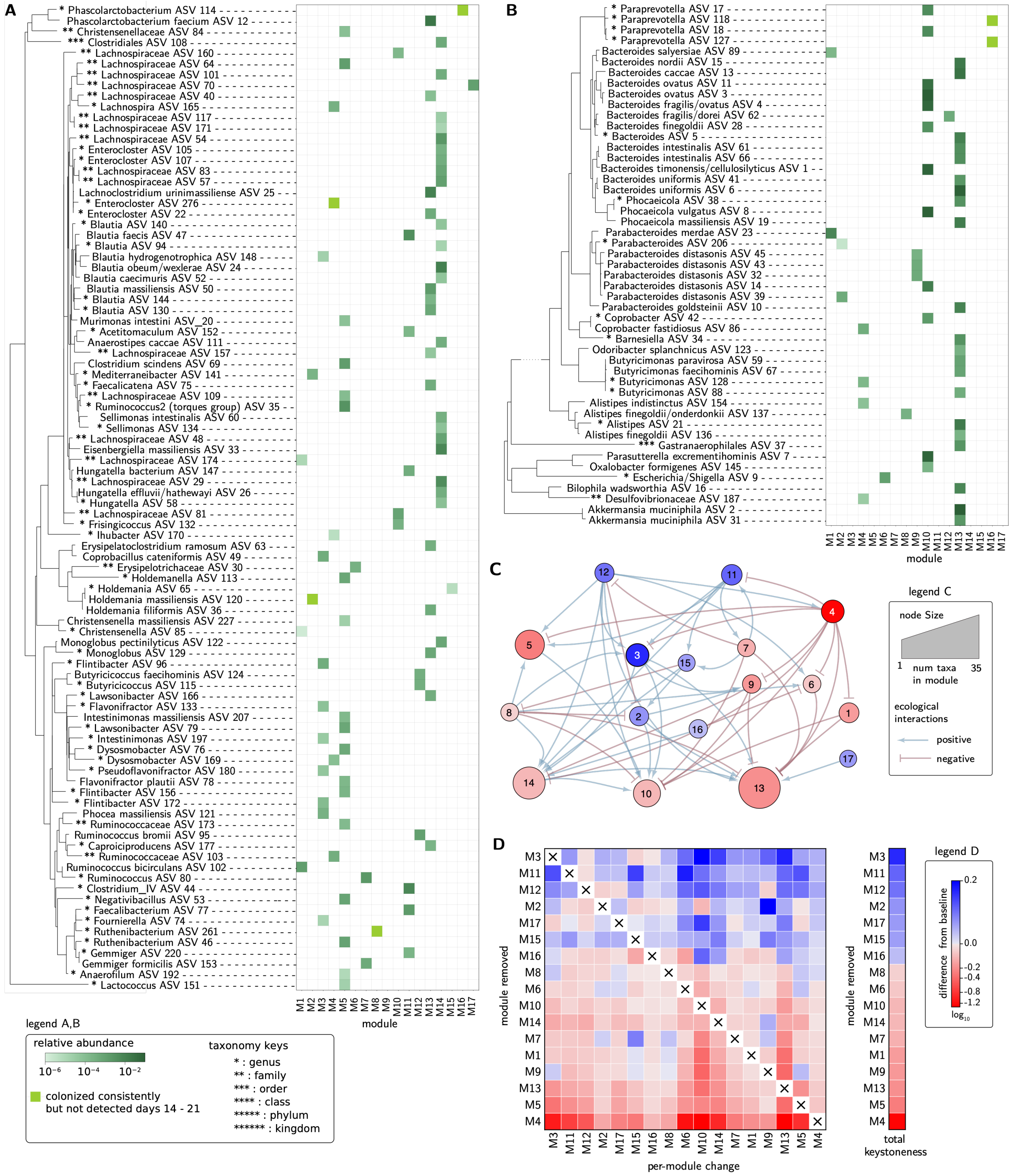


Figure 5: MDSINE2 infers modular representations of complex microbiome dynamics. Our method automatically learns modules of ASVs based on similarity of their dynamic interactions and responses to perturbations. Results are split into: **(A)** Gram-positive, and **(B)** Gram-negative ASVs, for display purposes. **(A2,B2)** Phylogenetic tree of ASVs. **(A3,B3)** Module memberships. Intensity of color in the grid indicates abundance post- colonization and prior to perturbations (average over Days 14 to Day 21). **(C)** Inferred module interaction network displaying only edges with BF > 100 (*decisive* evidence). Size of nodes are proportional to the number of ASVs in the module and color of nodes corresponds to module keystoneness. **(D)** Keystoneness analysis measures the relative importance of a module in maintaining the steady state abundance of the community. A positive keystoneness for a module means its removal from the community reduces the steady state abundance of its community members, and a negative keystoneness means that the module's removal results in an increase in the abundance of the other community members. A module's impact on the other community members was computed both at the module level and across all modules (total keystoneness).

MDSINE2 learned a compact set of interaction modules that revealed keystone sets of species and stability characteristics of a complex gut microbial ecosystem

On our benchmarking dataset, MDSINE2 discovered 17 interaction modules (Figure 5, Supplemental Table 8), ranging in size from 1 to 35 taxa, and connected through 56 interactions predicted with *decisive evidence* (Bayes Factor [BF] ≥ 100, Figure 5C)[32]. This represents a 97% reduction in interaction parameters over MDSINE2^{-M} (2179 edges predicted with *decisive evidence* [Supplemental Figure 5]), with nearly comparable forecasting performance as described in the previous section. As one measure of the biological relevance of modules, we evaluated the relatedness of taxa within modules compared to random assortments of taxa. Using phylogenetic distance as a metric, we found that the average within module distance was significantly lower (p=0.0135) than chance (Supplemental Figure 3). From the taxonomic perspective, five modules were significantly enriched at the Family level with two at the Order, two at the Class, and four at the Phylum levels, including expected segregation of modules along Gram positive and negative Phyla, e.g., Bacteroidetes and Firmicutes (Supplemental Figure 4)[32]. These results suggest that the modules, which are wholly learned from data measuring the dynamic behavior of the microbe, also significantly reflect phylogenetic signal.

To evaluate quantitatively the relative importance of interaction modules in the ecosystem, we performed a module-level keystoneness analysis[1] (Figure 5D, and Methods §4.6). Keystone taxa were originally defined as those taxa that are fundamental to the integrity of the ecological community, without whom the community would collapse[34], and have been suggested as drivers of microbial community structure and function[35,36]. Here, we extend the concept to groups of taxa (modules), to aid in interpretability, and also generalize to a quantitative measure of *keystoneness* with both positive and negative values. Positive keystone modules ("promoters") are those that when removed result in a reduction in the microbial abundances of the other members of the ecosystem; negative keystone modules ("suppressors") are those that when removed result in increases of abundances of the other members of the ecosystem. The magnitude of the keystoneness measure thus represents the degree of community wide disruption (in terms of microbial abundance change, with the removal of the module).

For our cohort, the top positive and negative keystoneness modules were M3 and M4 respectively. Investigating their role through the ecological network, we see that all the outgoing edges of M3 are promoting, while all the outgoing edges of M4 are repressive, suggesting the different ecological roles that these modules play in the network. M3 is enriched for the family Ruminocococaceae. Promoting M3 are two other positive keystoneness modules M11 and M12, each containing taxa capable of degrading resistant starches (*Ruminococcus bromii* ASV95, *Gemmiger* ASV220) and others with butyrate production capabilities (*Faecalibacterium* ASV77, *Butyricicoccus faecihominis* ASV124, *Butyricicoccus* ASV115). Downstream modules being promoted by M3 of note are M10 (enriched for Bacteroidaceae), M13 (enriched for Bacteroidetes and the largest module in the network), and M14 (enriched for Lachnospiraceae). One explanation for this structure, consistent with known biology, is that the positive keystoneness modules are connected in a cross-feeding chain beginning with specialized starch degrading taxa that ultimately support the more abundant generalist taxa (e.g., Bacteroidaceae). In contrast to the this specialist-to-generalist structure, the module with the highest negative keystoneness, M4, contains a diverse group of taxa and also suppresses multiple modules in the network, including M3 and M11, the top two positive keystoneness modules, as well as the primarily gram-negative modules M10 and M13. An annotated module network can be found in Supplemental Figure 6.

To assess the overall robustness of our inferred ecosystem to external perturbations we performed a formal stability analysis. Informally, stable systems are those systems that when undergoing a perturbation do not deviate far from their equilibrium state, and when that perturbation is removed, return to their original equilibrium state (asymptotic stability). Healthy gut microbiomes have many empirical features of a stable dynamical system, such as maintaining an equilibrium taxa abundance profile over long periods of time (years), and a rapid return to equilibrium (within days) after perturbations are removed (e.g. antibiotics, diet change, etc.)[33,37-39]. MDSINE2 learns an ecosystem-scale model of dynamics, and we can thus perform formal analyses of the underlying model parameters and equations to gauge stability. To compute stability, we used the fact that for gLV dynamics, a necessary condition for asymptotic stability is that the eigenvalues of the pairwise interactions matrix have all negative real parts (Supplemental Text §3)[40]. We note that prior studies have performed similar analyses, but for much smaller models trained on only tens of taxa[1,7,13].

Overall, we found that the dynamics of the model inferred were 80% likely to be stable (Figure 6B). This is consistent with prior work on smaller microbial networks suggesting that dynamics inferred from a healthy microbiome are indeed likely to be stable[7]. To assess how likely this scenario is given a network of the same size, we created a randomized network (random rewiring) that maintained a core topological feature of the network (the in-degree for each node was maintained) but that had the edges permuted, meaning that the number of interactions incoming to each node is maintained (Methods §4.8). With this random rewiring the system had a 0% probability of being stable (Figure 6B, Methods §4.7, Supplemental Text §3). These results indicate that the model MDSINE2 inferred from a healthy, complex gut microbiome is far more stable than would be expected by chance and highlights our method's ability to perform dynamical systems analyses at ecosystems-scale.

We next sought to identify features of the ecological network inferred by MDSINE2 that could explain the remarkable stability of healthy gut microbiomes. 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

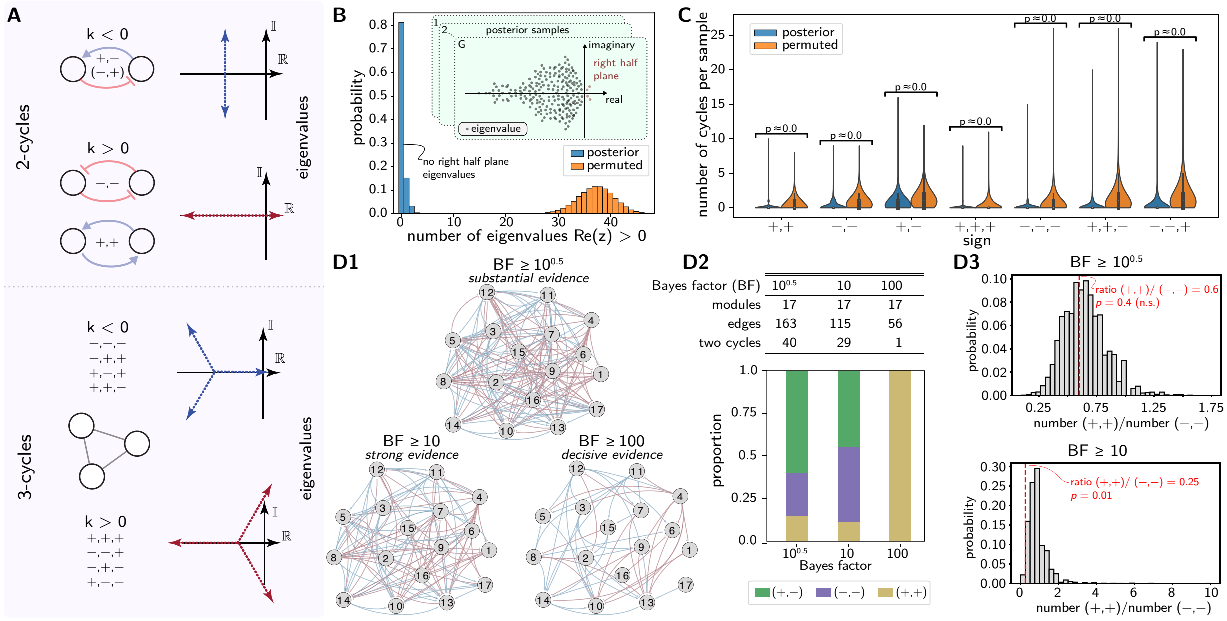


Figure 6: Analytical stability analysis of microbiomes demonstrates significantly more stable dynamics than expected by chance and reveals cycle feedback motifs. **(A)** two and three-cycle motifs and their corresponding eigenvalue asymptotes for increased interaction strength. Only the two-cycle with negative feedback can remain stable for all feedback gains (top motif). **(B)** As a measure of stability, we computed the number of right half plane eigenvalues for each posterior sample of our model trained on the microbiome data, and for the null model ("permuted"). Our model had an 80% probability of having no unstable eigenvalues. **(C)** Distribution of cycle motifs over posterior samples. Analyses showing **(D1)** networks derived with three different levels of confidence for including edges, **(D2)** descriptive statistics for the networks, and **(D3)** statical test for significance of mutualism to competition ratio (MCR [(+,+)/(−,−)]), which was significant for the network constructed from edges with *strong* evidence *p*=0.01.

well-characterized for linear and gLV dynamical systems[6]: mutualism (+,+) and competition (−,−) are destabilizing, and parasitism (+,−) is stabilizing. For length three cycles and higher, more complex ecological interactions arise, and any sign combination is potentially destabilizing[42] (Figure 6A). For all cycle lengths analyzed, we found that MDSINE2's inferred model of dynamics had a significantly lower number of cycles than expected by chance (Figure 6C). Of note, the most prevalent network motif in the inferred model was the always stable negative feedback 2-cycle (+,−); however, this motif was also present significantly less frequently than expected by chance.

We next sought to understand the influence of uncertainty in the inferred network structure itself on stability estimates, given that ecological networks of the microbiome are inferred from relatively limited and noisy data. To gain insight into this phenomenon, we evaluated networks at different levels of evidence for edges, *substantial* [$\mathrm{BF} \geq 10^{0.5}$], *strong* [$\mathrm{BF} \geq 10$], and *decisive* [$\mathrm{BF} \geq 100$] evidence (Figure 6D1, Supplemental Text §G). As the evidence threshold for an edge being included in the model was decreased, the number of edges in the network increased from 56 to 163 (Figure 6D2). As the number of edges increased, there was also an increase in the number of (two) cycles, as expected. Interestingly, for networks with more edges, the number of parasitism (+,−) cycles increased disproportionally among two cycles present, consistent with the property that stability becomes less likely the denser a network becomes (the more edges there are for a fixed number of nodes)[5,43,44], unless the cycles in the network are only parasitism (Figure 6A). Previous work has also examined the role of mutualism and competition cycles, hypothesizing that for healthy ecosystems, the mutualism to competition ratio (MCR [(+,+)/(−,−)]) would be less than one[45], and demonstrating this phenomenon on networks inferred on small microbial ecosystems[7]. Our results provide support for this hypothesis on the complex, healthy gut microbiome, showing the that the mutualism to competition ratio was significantly lower than chance on networks with strong evidence for the existence of edges (Figure 6D3).

Discussion

MDSINE2 is a fully Bayesian method that propagates uncertainty throughout the model[46], including explicit models of measurement noise in both relative abundance and total bacterial concentration, which we calibrate using physical and technical replicates. This framework allows us to achieve unprecedented performance improvements over prior methods, including our own, which did not fully propagate measurement uncertainty. Another important component of our model 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[47-49], but relatively underexploited in the microbiome field[19,20].

Dynamical systems inference requires data with sufficient richness of information. Measurements taken when trajectories are significantly deviating from steady state, e.g., transients, are critical for efficient inference[18]. To generate sufficiently rich data, we performed multiple strong perturbations in a cohort of gnotobiotic mice serially over time. This experimental design allowed us to perform temporally dense sampling with biological replicates (effectively multiple initial conditions), coupled with strong perturbations. Thus, our time-series data, specifically generated for efficient dynamical systems inference, is expected to be a valuable resource to any researcher studying ecological dynamics.

Dynamical systems analyses can offer insights into how ecosystems behave under physiologically relevant perturbations, and an understanding of these dynamics can then in turn be used to develop strategies to control the ecosystem[50-52]. Our analyses on longitudinal data from a cohort of mice colonized with a microbiome from a healthy human donor identified specific (e.g., putative cross-feeding chains beginning with resistant starch degrading organisms) and general (e.g., topological motifs) properties in the ecological network that promote stability. These findings, and the overall analysis methodology we have introduced, provide a foundation for designing interventions to increase the stability of dysbiotic microbiomes.

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[58]; 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 synthetic data and a new resource of densely sampled microbiome time-series data, from "humanized" gnotobiotic mice, that our approach outperforms other methods when forecasting microbiome dynamics or predicting ground truth dynamics. 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

1. MDSINE2 Model

1.1 Overview

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

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

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; $\boldsymbol{z}_{p,\mathbf{c}_i}^{(\gamma)}$ 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 $\boldsymbol{b}_{\mathbf{c}_i,\mathbf{c}_j}$, with $\boldsymbol{z}_{\mathbf{c}_i,\mathbf{c}_j}^{(b)}$ 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 support efficient inference, we use a first-order discretization (see Supplemental Text) to yield the discrete-time latent trajectories:

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

where

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

Here, $\Delta_{s,k} = t_{s,k+1} - t_{s,k}$, the difference between adjacent time-points for the time-series s. Below we give additional details on the model, including prior probability distributions on variables; for complete mathematical and algorithmic details, see Supplemental Text.

1.2 Interaction Modules

We employ a Dirichlet Process (DP) prior[59] 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[60]. 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[61]. Our formulation allows us to marginalize out the interaction and perturbation parameters during inference, which greatly increases efficiency[59]. See Supplemental Text for complete details.

1.3 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 Text for full details.

1.4 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[30]

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

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

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

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:

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

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 Text for complete mathematical details of the measurement model and inference procedure.

1.5 Software

MDSINE2 was implemented in Python 3.7 using the Numpy[62], Scipy[63], Numba[64], Matplotlib[65], and Seaborn[66] 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 objects, 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[67] 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 tutorials. The tutorials (Google Colab) can be accessed from https://github.com/gerberlab/MDSINE2_Paper

2. Gnotobiotic Experiments and Microbiome Data Generation

2.1 Mouse Experiments

A cohort of four male C57Bl/6 germfree mice was 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.[25] The mice were given a Fecal Microbiota Transplant (FMT) from a healthy human stool donor 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.

2.2 DNA Extraction, 16S rRNA Amplicon Sequencing and qPCR

For DNA extraction, all samples were processed using the standard protocol[68] 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[69]: 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.

3. Bioinformatics

3.1 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[28]. 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 1088 ASVs in total.

3.2 Phylogenetic Placement of Sequences

We performed phylogenetic placement of consensus ASVs 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 [70]. We performed multiple alignment of the sequences using the RDP's web-hosted alignment tool with default parameters [71]. To facilitate a good multiple alignment, we filtered out sequences with insertions seen in ≤ 3 other sequences. A reference tree was constructed using FastTree[72] 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[73]. ASV 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[74].

3.3 Fold Change Analysis

Fold change analysis was performed using DESeq2[29] 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/analysis/deseq. The fold changes were calculated using the default Wald test in the software. The fold changes during the perturbations were calculated 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 $(52,52.5,53,54)\ to$ $(46,47,50,50.5)$.

4. MDSINE2 analyses

4.1 Model Inference

Day 0 and 0.5 samples were excluded from inferences, due to their very low overall bacterial concentrations. First, using replicate data, we performed a fit for the negative-binomial model resulting in $d_0 = 4.2 \times 10^{-8}$, $d_1 = 6.05 \times 10^{-2}$. Using these hyperparameters, inference was performed using ten seeds configured to learn modules. For each seed, models were inferred using 10,000 MCMC iterations after 5,000 burn-in steps. Then, inference was performed on a single seed in fixed-cluster mode, where the consensus modules were derived using the concatenated outputs of the ten seeds. To assess convergence of each Markov chain, we used the $\hat{R}$ statistic[46] and confirmed values of $\hat{R}$<1.05 for the concentration parameter, the growth rates and the process variance, to indicate sufficient mixing. Inference took ~26.5 hours (6.3 sec/iteration) on an Intel Xeon E5-2697V2 (2.70 GHz base frequency, 3.50 GHz max frequency) with an allocation of one core and 8 Gb of RAM separately for each task. Full details on inference are given in the Supplemental Text.

4.2 Benchmarking

We used implementations of the comparator methods provided in https://github.com/tyjo/clv. Following Joseph et al. (2020)[15], we trained the models using elastic-net regression, and in addition, we trained gLV using ridge regression to provide comparisons to earlier work [1,13]. 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)[15]. 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 ASVs in the hold-out subjects.

4.3 Synthetic Data

Benchmarking of models with synthetic data in Figure 2 followed the data generation procedure used to benchmark MDSINE in the original Bucci et al. (2016)[1] manuscript. Briefly, gLV dynamics were simulated for a community of ten taxa with ten biological replicates for a total of 30 days each. Nine taxa were simulated in the system at day 0 and the challenge taxon was introduced on day ten. Measurements were assumed to be daily, with qPCR measurements simulated from a Lognormal Distribution and reads were simulated from a Dirichlet-Multinomial Distribution.

4.4 Consensus Module Construction and Fixed Module Inference

Consensus modules were constructed by performing agglomerative clustering on the co-clustering probability matrix where the number of clusters was the median number of modules over the posterior. See[19,20] for additional details.

4.5 Taxonomic Enrichment Analysis

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

4.6 Keystoneness

The keystoneness measure was 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, 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, are given by:

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

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

4.7 Stability

As a measure of stability, we computed the number of right half plane eigenvalues for the interaction matrix in each posterior sample. For a null model we generated in-degree preserving permutations of the interaction matrix (permutations are performed at the module level for each sample, see Methods §4.8). The number of right half plane eigenvalues for the interaction matrix was then determined by counting the number of eigenvalues whose real part was greater than zero. For a theoretical discussion on the use of the eigenvalues of the interaction matrix for determining stability, see Supplemental Text §3.1.

4.8 Network Null Model

To provide test of statistical significance for network topological features, we generated indegree-preserving permutations of the interaction matrix, performed by permuting the off-diagonal elements in each row, to serve as null distributions of the network topological features. With this permutation, the total number of edges in the network remains the same and the numbers of edges coming into each node remains the same, but the sources for the edges are uniformly randomly assigned with each permutation.

References

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

2 Gerber, G. K. The dynamic microbiome. *FEBS letters* **588**, 4131-4139 (2014).

3 Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nature Medicine* **24**, 392-400 (2018). https://doi.org:10.1038/nm.4517

4 Gonze, D., Coyte, K. Z., Lahti, L. & Faust, K. Microbial communities as dynamical systems. *Current opinion in microbiology* **44**, 41-49 (2018).

5 May, R. M. Will a large complex system be stable? *Nature* **238**, 413-414 (1972).

6 Goh, B. S. Global Stability in Many-Species Systems. *The American Naturalist* **111**, 135-143 (1977).

7 Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, competition, and stability. *Science* **350**, 663-666 (2015).

8 Rao, C. *et al.* Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature* **591**, 633-638 (2021). https://doi.org:10.1038/s41586-021-03241-8

9 Lee, T. I. *et al.* Transcriptional regulatory networks in Saccharomyces cerevisiae. *science* **298**, 799-804 (2002).

10 Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827 (2002).

11 Barabasi, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature reviews genetics* **5**, 101-113 (2004).

12 Franzosa, E. A. *et al.* Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology* **4**, 293-305 (2019). https://doi.org:10.1038/s41564-018-0306-4

13 Stein, R. R. *et al.* Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. *PLoS Comput Biol* **9** (2013).

14 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).

15 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 (2020). https://doi.org:10.1371/journal.pcbi.1007917

16 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).

17 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 (2012). https://doi.org:10.1038/ismej.2012.4

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

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

20 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 (2020). https://doi.org:10.1186/s13073-020-00758-x

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

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

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

24 Gibson, T. & Gerber, G. in *Proceedings of the 35th International Conference on Machine Learning* Vol. 80 1763-1772 (2018).

25 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 (2018). https://doi.org:10.1002/cpim.48

26 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).

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

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

29 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014). https://doi.org:10.1186/s13059-014-0550-8

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

31 Greenacre, M. Compositional data analysis. *Annual Review of Statistics and its Application* **8**, 271-299 (2021).

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

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

34 Paine, R. T. A note on trophic complexity and community stability. *The American Naturalist* **103**, 91-93 (1969).

35 Banerjee, S., Schlaeppi, K. & van der Heijden, M. G. Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology* **16**, 567-576 (2018).

36 Röttjers, L. & Faust, K. Can we predict keystones? *Nature Reviews Microbiology* **17**, 193-193 (2019). https://doi.org:10.1038/s41579-018-0132-y

37 Faith, J. J. *et al.* The Long-Term Stability of the Human Gut Microbiota. *Science* **341**, 1237439 (2013). https://doi.org:10.1126/science.1237439

38 Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220-230 (2012). https://doi.org:10.1038/nature11550

39 Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences* **108**, 4554-4561 (2011). https://doi.org:10.1073/pnas.1000087107

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 (2016). https://doi.org:10.1371/journal.pcbi.1004688

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 Allesina, S. & Tang, S. Stability criteria for complex ecosystems. *Nature* **483**, 205--208 (2012).

44 Sommers, H. J., Crisanti, A., Sompolinsky, H. & Stein, Y. Spectrum of large random asymmetric matrices. *Physical Review Letters* **60**, 1895--1898 (1988).

45 Palmer, J. D. & Foster, K. R. Bacterial species rarely work together. *Science* **376**, 581-582 (2022).

46 Gelman, A. *et al. Bayesian data analysis*. (Chapman and Hall/CRC, 2013).

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

48 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).

49 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 (2007). https://doi.org:10.1371/journal.pcbi.0030148

50 Gibson, T. E. in *American Control Conference* (IEEE, 2016).

51 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 Comput Biol* **12**, e1004688 (2016). https://doi.org:10.1371/journal.pcbi.1004688

52 Angulo, M. T., Moog, C. H. & Liu, Y.-Y. A theoretical framework for controlling complex microbial communities. *Nature Communications* **10**, 1045 (2019). https://doi.org:10.1038/s41467-019-08890-y

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 (2019). https://doi.org:doi:10.1128/mBio.02566-18

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

55 Joossens, M. *et al.* Dysbiosis of the faecal microbiota in patients with Crohn&#039;s disease and their unaffected relatives. *Gut* **60**, 631 (2011). https://doi.org:10.1136/gut.2010.223263

56 Kowalska-Duplaga, K. *et al.* Differences in the intestinal microbiome of healthy children and patients with newly diagnosed Crohn's disease. *Sci Rep* **9**, 18880 (2019). https://doi.org:10.1038/s41598-019-55290-9

57 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 (2018). https://doi.org:10.1186/s40168-018-0603-4

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

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

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

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

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

63 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods* **17**, 261-272 (2020).

64 Lam, S. K., Pitrou, A. & Seibert, S. in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC.* 1-6.

65 Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in science & engineering* **9**, 90-95 (2007).

66 Waskom, M. & team, S. d. mwaskom/seaborn. (2020). <https://doi.org/10.5281/zenodo.592845>.

67 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).

68 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 (2019). https://doi.org:10.1016/j.chom.2019.05.001

69 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).

70 Cole, J. R. *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**, D141-D145 (2009).

71 Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335-1337 (2009).

72 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* **5**, e9490 (2010).

73 Eddy, S. R. Accelerated profile HMM searches. *PLoS computational biology* **7**, e1002195 (2011).

74 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).

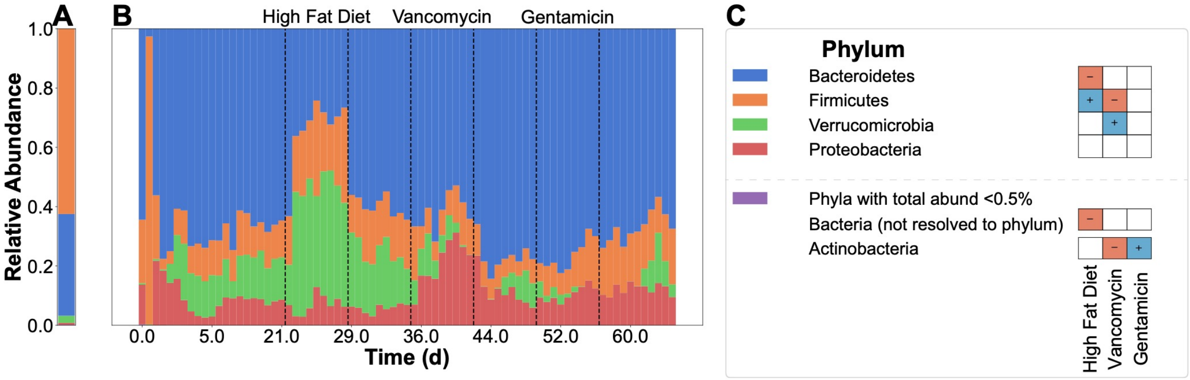
Contributions:

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.

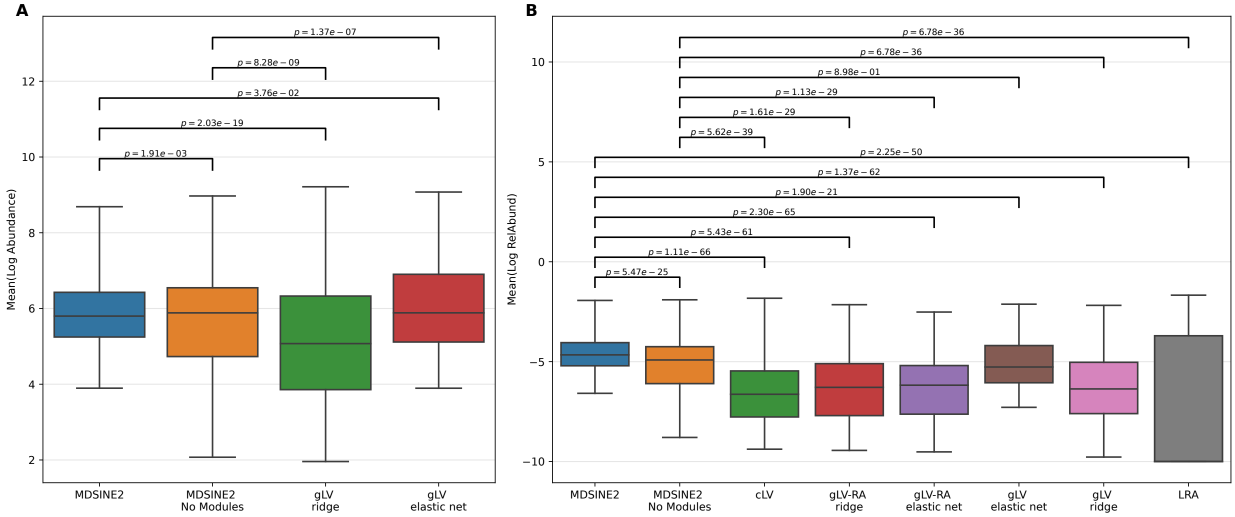
Competing Interests:
No industry support was provided for this study.
TEG: None.
YK: None.
SA: None.
DK: None.
ND: None.
RL: None.
BB: None.
JRA consults for Finch Therapeutics, BMS, Pfizer, Janssen, Morphic, Iterative Scopes, Artugen, Servatus, Pandion, Merck, Baccain and has research support from Merck. 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 R35GM141861 (Berger), NIH P30DK056338 (Bry), NSF MTM2 2025512 (Gerber), the BWH President's Scholar Award (Gerber)

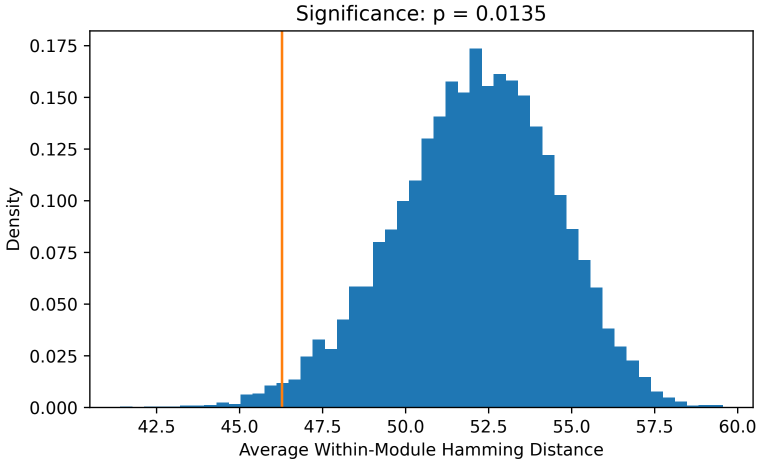
Software and Data Availability: First time users wanting to explore the model and the data should start by visiting the repo for this paper that reproduces figures and contains notebooks for exploring the model https://github.com/gerberlab/MDSINE2_Paper. The core MDSINE2 package is available at https://github.com/gerberlab/MDSINE2. All sequencing data for this study are available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA784519.



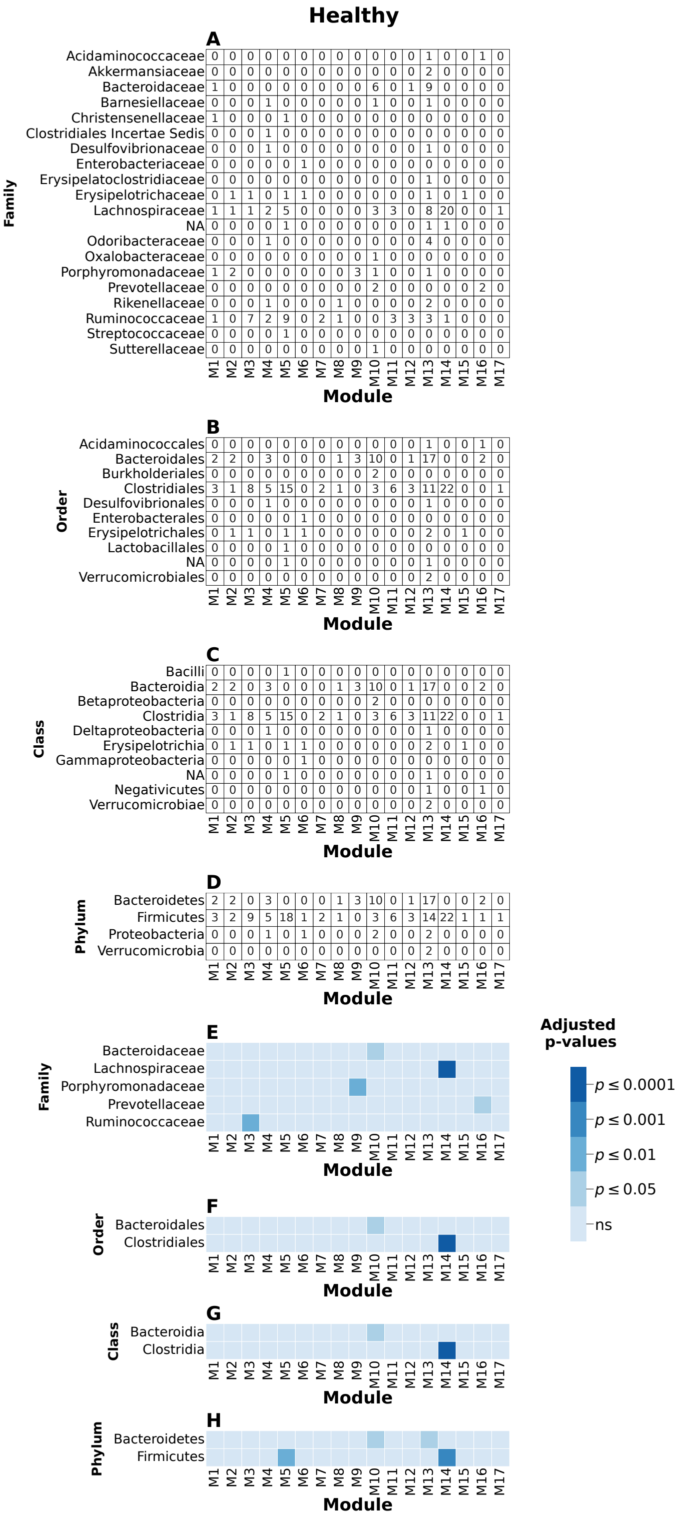
Supplemental Figure 1: Phylum level grouping of reads for the high-temporal resolution gnotobiotic mouse study. (A) Relative abundance of microbes in human donor sample. **(B)** Relative abundances of microbes in serial fecal samples, averaged over the biological replicates. **(C)** Legend and taxonomy for panels A and B along with significant differential abundances for each taxonomic group across the three perturbations (see Supplemental Table 1 for *p*-values).



Supplemental Figure 2: All models have relatively poor performance for the zero read taxa samples with errors on the order of the limit of detection. Results are for cross-fold validation of forecasts. Models were trained on three of four mice, then given the initial condition of the hold out mouse time-series, and then the inferred model was simulated forwards and the output was compared to the entire held out time-series. All tests are two-sided Wilcoxon signed-rank test (paired non-parametric). Non-significant tests are denoted with red bracket. Errors are calculated from forecasted abundance with respect to hold out data across all folds and for all taxa with zero reads within each sample. **(A)** Forecasting error for all zero taxa samples for models trained on concentrations. **(B)** Forecasting error for all zero taxa samples for models trained on relative abundances.



Supplemental Figure 3: Phylogenetic distance of modules. Average phylogenetic distance of modules shown by orange line, compared to random permutations of taxa in modules for null distribution in blue. Permutation test results in p=0.0135.



A

Family	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17
Acidaminococcaceae	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
Akkermansiaceae	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
Bacteroidaceae	1	0	0	0	0	0	0	0	0	6	0	1	9	0	0	0	0
Barnesiellaceae	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0
Christensenellaceae	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Clostridiales Incertae Sedis	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Desulfovibrionaceae	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
Enterobacteriaceae	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Erysipelatoclostridiaceae	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Erysipelotrichaceae	0	1	1	0	1	1	0	0	0	0	0	0	1	0	1	0	0
Lachnospiraceae	1	1	1	2	5	0	0	0	0	3	3	0	8	20	0	0	1
NA	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0
Odoribacteraceae	0	0	0	1	0	0	0	0	0	0	0	0	4	0	0	0	0
Oxalobacteraceae	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Porphyromonadaceae	1	2	0	0	0	0	0	0	3	1	0	0	1	0	0	0	0
Prevotellaceae	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0
Rikenellaceae	0	0	0	1	0	0	0	1	0	0	0	0	2	0	0	0	0
Ruminococcaceae	1	0	7	2	9	0	2	1	0	0	3	3	3	1	0	0	0
Streptococcaceae	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Sutterellaceae	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

B

Order	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17
Acidaminococcales	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
Bacteroidales	2	2	0	3	0	0	0	1	3	10	0	1	17	0	0	2	0
Burkholderiales	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Clostridiales	3	1	8	5	15	0	2	1	0	3	6	3	11	22	0	0	1
Desulfovibrionales	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
Enterobacterales	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Erysipelotrichales	0	1	1	0	1	1	0	0	0	0	0	0	2	0	1	0	0
Lactobacillales	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
NA	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Verrucomicrobiales	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0

C

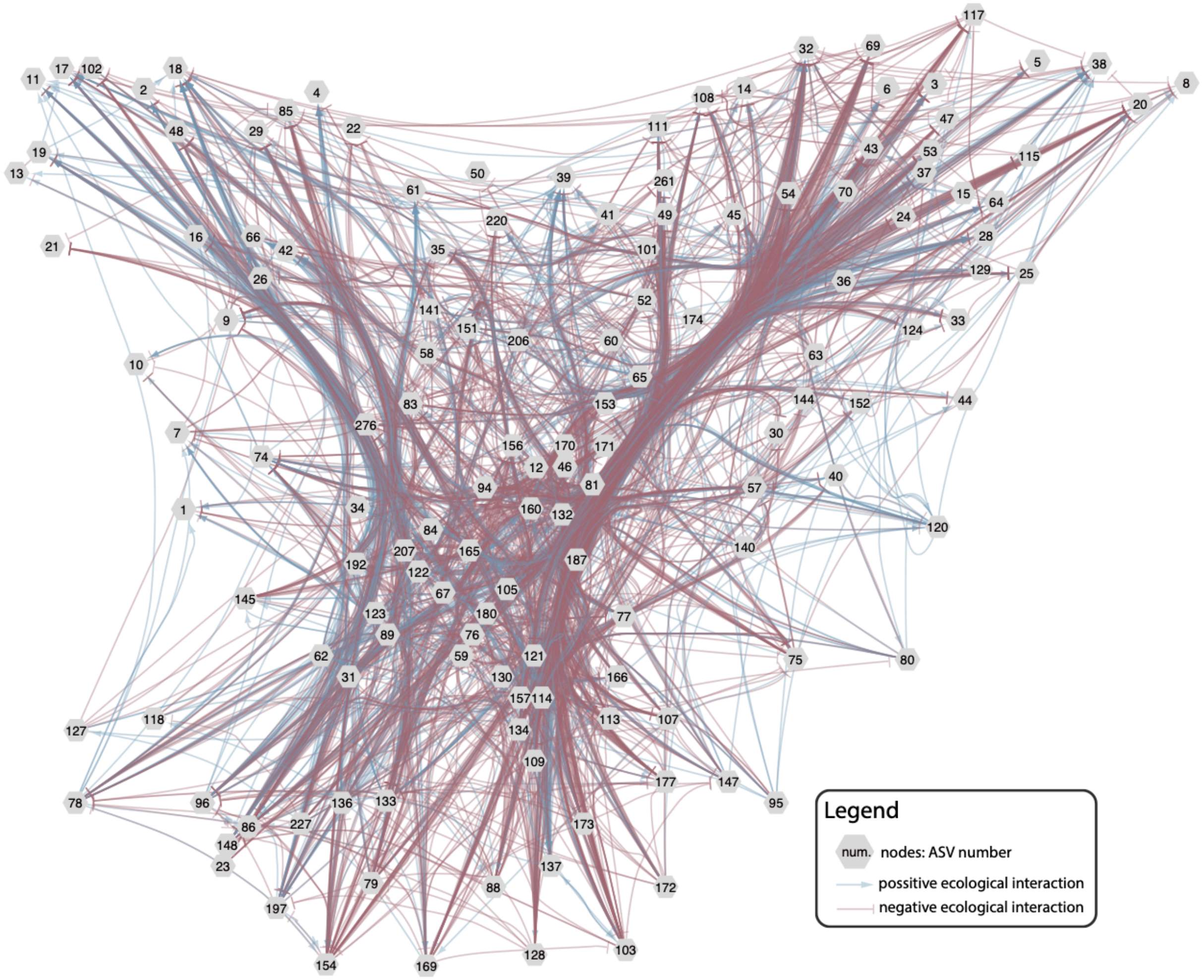
Class	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17
Bacilli	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Bacteroidia	2	2	0	3	0	0	0	1	3	10	0	1	17	0	0	2	0
Betaproteobacteria	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Clostridia	3	1	8	5	15	0	2	1	0	3	6	3	11	22	0	0	1
Deltaproteobacteria	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
Erysipelotrichia	0	1	1	0	1	1	0	0	0	0	0	0	2	0	1	0	0
Gammaproteobacteria	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
NA	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Negativicutes	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
Verrucomicrobiae	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0

D

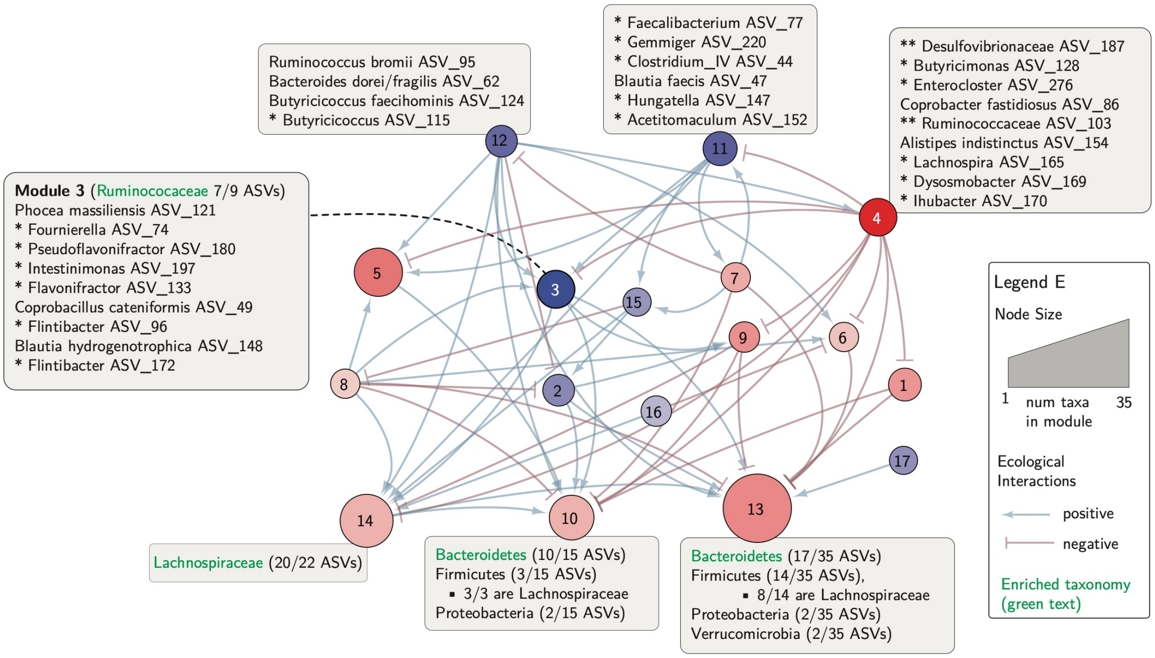
Phylum	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17
Bacteroidetes	2	2	0	3	0	0	0	1	3	10	0	1	17	0	0	2	0
Firmicutes	3	2	9	5	18	1	2	1	0	3	6	3	14	22	1	1	1
Proteobacteria	0	0	0	1	0	1	0	0	0	2	0	0	2	0	0	0	0
Verrucomicrobia	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0

NA: Taxonomy not resolved

Supplemental Figure 4: Taxonomic enrichment. Enrichment for the modules was performed at Phylum, Class, Order, and Family levels using the hypergeometric test, Methods §4.5.


117
32
69
5
38
8
11
17
102
18
2
4
108
14
6
3
20
85
48
29
22
111
47
43
53
19
13
115
50
61
39
261
37
54
70
15
64
41
49
45
220
24
28
21
16
66
42
35
101
129
25
26
36
52
141
174
33
9
151
206
124
60
58
63
10
65
83
153
144
152
44
276
30
7
156
170
171
74
12
46
81
40
94
57
34
1
160
132
120
84
140
207
165
187
192
122
105
145
67
123
180
77
89
76
62
59
121
75
80
31
130
166
157
114
127
118
134
113
107
109
177
147
78
96
136
133
95
86
227
173
148
23
137
79
88
172
197
103
128
154
169
Legend
num.
nodes: ASV number
possitive ecological interaction
negative ecological interaction

Supplemental Figure 5. Taxon-taxon interaction network for model trained without module inference. Interaction network displaying only edges with BF > 100 (*decisive* evidence).


* Faecalibacterium ASV_77
* Gemmiger ASV_220
* Clostridium_IV ASV_44
Blautia faecis ASV_47
* Hungatella ASV_147
* Acetitomaculum ASV_152
Ruminococcus bromii ASV_95
Bacteroides dorei/fragilis ASV_62
Butyricicoccus faecihominis ASV_124
* Butyricicoccus ASV_115
** Desulfovibrionaceae ASV_187
* Butyricimonas ASV_128
* Enterocloster ASV_276
Coprobacter fastidiosus ASV_86
** Ruminococcaceae ASV_103
Alistipes indistinctus ASV_154
* Lachnospira ASV_165
* Dysosmobacter ASV_169
* Ihubacter ASV_170
Module 3 (Ruminococaceae 7/9 ASVs)
Phocea massiliensis ASV_121
* Fournierella ASV_74
* Pseudoflavonifractor ASV_180
* Intestinimonas ASV_197
* Flavonifractor ASV_133
Coprobacillus cateniformis ASV_49
* Flintibacter ASV_96
Blautia hydrogenotrophica ASV_148
* Flintibacter ASV_172
12
11
4
5
3
15
7
9
6
8
2
16
1
17
14
10
13
Legend E
Node Size
1
num taxa in module
35
Ecological Interactions
positive
negative
Enriched taxonomy (green text)
Lachnospiraceae (20/22 ASVs)
Bacteroidetes (10/15 ASVs)
Firmicutes (3/15 ASVs)
▪ 3/3 are Lachnospiraceae
Proteobacteria (2/15 ASVs)
Bacteroidetes (17/35 ASVs)
Firmicutes (14/35 ASVs),
▪ 8/14 are Lachnospiraceae
Proteobacteria (2/35 ASVs)
Verrucomicrobia (2/35 ASVs)

Supplemental Figure 6: Annotated modules from Figure 5C. Taxa from specific modules denoted by ASVs along with taxonomic breakdown at higher levels. Enriched taxonomies in green. Only to aid in discussion, not intended to be comprehensive.

Growth Rate						
Method A	Method B	Median A	Median B	Median A minus B	p	BHCorrectedp
MDSINE2	MDSINE1	0.120675231	0.15621351	-0.034542389	0.01953125	0.01953125
MDSINE2	glv-ridge	0.120675231	0.2810688	-0.160663543	0.001953125	0.002441406
MDSINE2	glv-elastic_net	0.120675231	0.28258131	-0.160474627	0.001953125	0.002441406
MDSINE1	glv-ridge	0.156213507	0.2810688	-0.129547605	0.001953125	0.002441406
MDSINE1	glv-elastic_net	0.156213507	0.28258131	-0.131218719	0.001953125	0.002441406
Interaction Strength						
Method A	Method B	Median A	Median B	Median A minus B	p	BHCorrectedp
MDSINE2	MDSINE1	7.46E-06	2.58E-05	-1.81E-05	0.001953125	0.001953125
MDSINE2	glv-ridge	7.46E-06	1.74E-05	-1.04E-05	0.001953125	0.001953125
MDSINE2	glv-elastic_net	7.46E-06	1.92E-05	-1.16E-05	0.001953125	0.001953125
MDSINE1	glv-ridge	2.58E-05	1.74E-05	9.06E-06	0.001953125	0.001953125
MDSINE1	glv-elastic_net	2.58E-05	1.92E-05	6.79E-06	0.001953125	0.001953125
Network Topology						
Method A	Method B	Median A	Median B	Median A minus B	p	BHCorrectedp
MDSINE2	MDSINE1	0.944251543	0.78819444	0.155864198	0.001953125	0.002441406
MDSINE2	glv-ridge	0.944251543	0.61824846	0.319637346	0.001953125	0.002441406
MDSINE2	glv-elastic_net	0.944251543	0.62056327	0.322916667	0.001953125	0.002441406
MDSINE1	glv-ridge	0.788194444	0.61824846	0.158564815	0.001953125	0.002441406
MDSINE1	glv-elastic_net	0.788194444	0.62056327	0.172260802	0.00390625	0.00390625

Supplemental Table 1: *p*-values for Figure 2.

	Perturbation 1						Perturbation 2						Perturbation 2					
	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Actinobacteria	7.7884928	1.504579628	0.94659	1.58948	0.11195	0.13994	7.7884928	-4.566340275	1.41752	-3.2214	0.00128	0.00319	7.7884928	4.628977762	1.42891	3.23952	0.0012	0.00599
Bacteroidetes	140423.6	-1.257655422	0.18352	-6.8529	7.24E-12	3.62E-11	140423.6	0.078989709	0.18351	0.43044	0.66687	0.83109	140423.6	0.005878835	0.1835	0.03204	0.97444	0.97444
Firmicutes	26401.99	0.594199148	0.23053	2.57752	0.00995	0.01659	26401.99	-0.66386383	0.23056	-2.8793	0.00399	0.00664	26401.99	0.522466401	0.23062	2.26552	0.02348	0.0587
Proteobacteria	23077.939	-0.098447524	0.27958	-0.3521	0.72474	0.72474	23077.939	1.943359082	0.27955	6.95171	3.61E-12	1.80E-11	23077.939	0.314675587	0.27957	1.12555	0.26036	0.41225
unknown	7.5082311	-5.5052887	1.28928	-4.27	1.95E-05	4.89E-05	7.5082311	0.174752093	0.81927	0.2133	0.83109	0.83109	7.5082311	-1.328392602	1.36313	-0.9745	0.3298	0.41225
Verrucomicrobia	27877.96	1.505666498	0.66245	2.27286	NA	NA	27877.96	-0.899346853	0.66248	-1.3575	NA	NA	27877.96	-1.78773361	0.66256	-2.6982	NA	NA

Supplemental Table 2: *p*-values for Supplemental Figure 1C.

	Perturbation 1						Perturbation 2						Perturbation 3					
	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Actinobacteria Coriobacteriia Coriobacteriales Coriobacteriaceae	4.2978873	0.903088847	1.92569	0.46897	0.63909	0.71578	4.2978873	-2.662582978	2.26878	-1.1736	0.24056	0.35027	4.2978873	0	2.3937	0	1	1
Bacteroidetes Bacteroidia Bacteroidales Bacteroidaceae	110643.9	-1.85921061	0.24788	-7.5004	6.36E-14	8.91E-13	110643.9	0.285016827	0.24787	1.14987	0.2502	0.35027	110643.9	-0.413812617	0.24787	-1.6695	0.09502	0.22172
Bacteroidetes Bacteroidia Bacteroidales Barnesiellaceae	1010.4248	-0.073385273	0.56661	-0.1295	0.89695	0.89695	1010.4248	-0.241386158	0.56699	-0.4257	0.6703	0.79707	1010.4248	0.300835224	0.56677	0.53078	0.59557	0.7559
Bacteroidetes Bacteroidia Bacteroidales Odoribacteraceae	525.93767	-0.597595855	0.4239	-1.4097	0.15861	0.21548	525.93767	0.202321106	0.42498	0.47607	0.63402	0.79707	525.93767	-2.08653854	0.42527	-4.9064	9.28E-07	1.30E-05
Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae	18599.491	-0.197695583	0.30778	-0.6423	0.52066	0.60744	18599.491	0.115534136	0.30775	0.37542	0.70735	0.79707	18599.491	0.238704692	0.30765	0.7759	0.43781	0.63811
Bacteroidetes Bacteroidia Bacteroidales Prevotellaceae	4293.0484	-11.4020524	1.13258	-10.067	7.71E-24	2.16E-22	4293.0484	-1.344895682	0.79015	-1.7021	0.08874	0.16565	4293.0484	2.486991151	0.79015	3.14748	0.00165	0.00922
Bacteroidetes Bacteroidia Bacteroidales Rikenellaceae	1779.3231	-1.293517088	0.71548	-1.8079	0.07062	0.12359	1779.3231	-1.265911203	0.71571	-1.7687	0.07694	0.15387	1779.3231	-0.79870041	0.71583	-1.1158	0.26452	0.46291
Firmicutes Bacilli Lactobacillales Streptococcaceae	39.313297	6.234269894	1.95576	3.18764	0.00143	0.00342	39.313297	-0.708996435	2.40994	-0.2942	0.76861	0.79707	39.313297	0	2.40994	0	1	1
Firmicutes Clostridia Clostridiales Catabacteriaceae	3.8277388	2.946567175	1.41271	2.08576	0.037	0.06907	3.8277388	-2.154172498	1.85633	-1.1604	0.24587	0.35027	3.8277388	-1.924453066	1.47935	-1.3009	0.1933	0.36082
Firmicutes Clostridia Clostridiales Christensenellaceae	185.91922	4.961099564	1.43799	3.45003	0.00056	0.00174	185.91922	-4.057708138	1.59792	-2.5394	0.01111	0.02827	185.91922	7.001891908	1.4019	4.99456	5.90E-07	1.30E-05
Firmicutes Clostridia Clostridiales Clostridiales_Incertae_Sedis_XIII	40.208404	1.405083309	1.00386	1.39969	0.16161	0.21548	40.208404	2.989501098	0.99649	3.00002	0.0027	0.00756	40.208404	-0.450584708	0.98668	-0.4567	0.64791	0.7559
Firmicutes Clostridia Clostridiales Eubacteriaceae	32.663345	1.159875629	1.11366	1.04149	0.29765	0.37492	32.663345	0.362878153	1.13619	0.31938	0.74944	0.79707	32.663345	2.262708076	1.11466	2.02995	0.04236	0.13179
Firmicutes Clostridia Clostridiales Lachnospiraceae	12436.111	0.089495779	0.45337	0.1974	0.84351	0.89695	12436.111	-4.22220286	0.45374	-9.3054	1.34E-20	1.87E-19	12436.111	0.926861634	0.4535	2.04381	0.04097	0.13179
Firmicutes Clostridia Clostridiales Peptostreptococcaceae	18.354898	7.376368918	2.33592	3.15781	0.00159	0.00342	18.354898	-0.109824236	2.59313	-0.0424	0.96622	0.96622	18.354898	-1.923669375	2.57937	-0.7458	0.45579	0.63811
Firmicutes Clostridia Clostridiales Ruminococcaceae	2713.7942	0.535708433	0.52547	1.01949	0.30797	0.37492	2713.7942	-2.428210578	0.52836	-4.5957	4.31E-06	3.02E-05	2713.7942	1.127013968	0.52629	2.14143	0.03224	0.13179
Firmicutes Clostridia Clostridiales unknown	118.29812	-0.997342921	0.61773	-1.6145	0.10641	0.15682	118.29812	-1.991420293	0.63526	-3.1348	0.00172	0.00535	118.29812	1.200439427	0.62291	1.92713	0.05396	0.1511
Firmicutes Clostridia unknown unknown	19.568408	-0.313740001	2.13903	-0.1467	0.88339	0.89695	19.568408	8.21217522	1.88274	4.36182	1.29E-05	7.22E-05	19.568408	-0.737921784	1.59241	-0.4634	0.64308	0.7559
Firmicutes Erysipelotrichia Erysipelotrichales Erysipelatoclostridiaceae	134.26542	1.420908483	0.86942	1.63432	0.10219	0.15682	134.26542	1.193947864	0.86591	1.37884	0.16794	0.27661	134.26542	3.287211139	0.8755	3.75465	0.00017	0.00162
Firmicutes Erysipelotrichia Erysipelotrichales Erysipelotrichaceae	2947.8726	2.285501996	0.55378	4.1271	3.67E-05	0.00017	2947.8726	4.289910059	0.55398	7.74383	9.65E-15	9.00E-14	2947.8726	-1.144383363	0.55331	-2.0683	0.03862	0.13179
Firmicutes Negativicutes Acidaminococcales Acidaminococcaceae	6164.4171	-0.93158979	0.36113	-2.5797	0.00989	0.01978	6164.4171	1.292824596	0.36106	3.58062	0.00034	0.00137	6164.4171	-0.12473667	0.36114	-0.3454	0.7298	0.81737
Firmicutes unknown unknown unknown	144.62473	5.228760889	1.65156	3.16596	0.00155	0.00342	144.62473	-7.955462313	1.94052	-4.0997	4.14E-05	0.00019	144.62473	0.114738234	1.77528	0.06463	0.94847	1
Proteobacteria Betaproteobacteria Burkholderiales Oxalobacteraceae	31.702053	-6.81404615	1.86182	-3.6599	0.00025	0.00088	31.702053	-3.14996853	1.68576	-1.8686	0.06168	0.13285	31.702053	2.192217096	1.60026	1.36991	0.17071	0.34143
Proteobacteria Betaproteobacteria Burkholderiales Sutterellaceae	8459.9419	-3.813127547	0.5152	-7.4012	1.35E-13	1.26E-12	8459.9419	-1.819906334	0.51491	-3.5344	0.00041	0.00143	8459.9419	0.838489645	0.51486	1.62858	0.1034	0.22271
Proteobacteria Deltaproteobacteria Desulfovibrionales Desulfovibrionaceae	3475.4961	2.852256507	0.4654	6.12862	8.86E-10	4.96E-09	3475.4961	0.988597209	0.46583	2.12225	0.03382	0.07891	3475.4961	0.834310581	0.46553	1.79217	0.07311	0.18609
Proteobacteria Gammaproteobacteria Enterobacterales Enterobacteriaceae	11939.646	7.796060276	1.20802	6.4536	1.09E-10	7.65E-10	11939.646	12.01969571	1.05196	11.426	3.10E-30	8.68E-29	11939.646	-0.916517738	1.0321	-0.888	0.37454	0.58261
Proteobacteria unknown unknown unknown	460.76052	-5.25903792	1.64809	-3.191	0.00142	0.00342	460.76052	1.163890718	1.6391	0.71008	0.47766	0.63687	460.76052	-0.934051978	1.63956	-0.5697	0.56888	0.7559
unknown unknown unknown unknown	8.954586	-5.903490037	1.48688	-3.9704	7.18E-05	0.00029	8.954586	0.374742696	1.1009	0.3404	0.73356	0.79707	8.954586	-1.513309229	1.5715	-0.963	0.33556	0.55269
Verrucomicrobia Verrucomicrobiae Verrucomicrobiales Akkermansiaceae	25674.698	1.157846979	0.70122	1.6512	0.0987	0.15682	25674.698	-1.016134776	0.70124	-1.449	0.14732	0.25782	25674.698	-2.222534889	0.70133	-3.169	0.00153	0.00922

Supplemental Table 3: *p*-values for Figure 3B.

Method A	Method B	Median A	Median B	Median A minus B	p	BHCorrectedp
MDSINE2	MDSINE2 (No Modules)	1.048985996	0.893952453	0.108802683	1.60E-14	1.60E-14
MDSINE2	gLV (ridge)	1.048985996	2.019284462	-0.836600299	1.97E-60	5.90E-60
MDSINE2	gLV (elastic net)	1.048985996	1.426345167	-0.318418115	1.97E-30	2.36E-30
MDSINE2 (No Modules)	gLV (ridge)	0.893952453	2.019284462	-0.944098552	3.04E-82	1.83E-81
MDSINE2 (No Modules)	gLV (elastic net)	0.893952453	1.426345167	-0.470104273	8.85E-58	1.77E-57
gLV (ridge)	gLV (elastic net)	2.019284462	1.426345167	0.373514595	3.09E-31	4.64E-31

Supplemental Table 4: *p*-values associated with Figure 4A.

Method A	Method B	Bin	Median A	Median B	Median A minus B	p	BHCorrectedp
MDSINE2	MDSINE2 (No Modules)	(5.0, 6.6]	0.957885324	1.012683899	-0.076074539	0.000244916	0.000267181
MDSINE2	gLV (ridge)	(5.0, 6.6]	0.957885324	1.459674664	-0.480065004	4.97E-25	1.36E-24
MDSINE2	gLV (elastic net)	(5.0, 6.6]	0.957885324	1.080218387	-0.15952346	2.21E-06	2.50E-06
MDSINE2 (No Modules)	gLV (ridge)	(5.0, 6.6]	1.012683899	1.459674664	-0.304563106	1.02E-16	1.53E-16
MDSINE2 (No Modules)	gLV (elastic net)	(5.0, 6.6]	1.012683899	1.080218387	-0.099090379	0.010300539	0.010842673
gLV (ridge)	gLV (elastic net)	(5.0, 6.6]	1.459674664	1.080218387	0.190720508	1.44E-13	1.96E-13
MDSINE2	MDSINE2 (No Modules)	(6.6, 7.0]	0.975838133	0.939439106	-0.007930717	0.834019116	0.834019116
MDSINE2	gLV (ridge)	(6.6, 7.0]	0.975838133	1.684145182	-0.637352929	1.92E-36	7.68E-36
MDSINE2	gLV (elastic net)	(6.6, 7.0]	0.975838133	1.268046327	-0.25056735	1.58E-11	1.97E-11
MDSINE2 (No Modules)	gLV (ridge)	(6.6, 7.0]	0.939439106	1.684145182	-0.623156068	1.70E-35	5.99E-35
MDSINE2 (No Modules)	gLV (elastic net)	(6.6, 7.0]	0.939439106	1.268046327	-0.300296498	5.90E-12	7.86E-12
gLV (ridge)	gLV (elastic net)	(6.6, 7.0]	1.684145182	1.268046327	0.267260144	8.05E-20	1.46E-19
MDSINE2	MDSINE2 (No Modules)	(7.0, 7.3]	1.010403313	0.962692177	-0.006232707	0.308834817	0.319484294
MDSINE2	gLV (ridge)	(7.0, 7.3]	1.010403313	1.877331449	-0.685043219	2.19E-44	1.20E-43
MDSINE2	gLV (elastic net)	(7.0, 7.3]	1.010403313	1.387608197	-0.309614975	3.11E-18	5.12E-18
MDSINE2 (No Modules)	gLV (ridge)	(7.0, 7.3]	0.962692177	1.877331449	-0.794427865	1.76E-46	1.17E-45
MDSINE2 (No Modules)	gLV (elastic net)	(7.0, 7.3]	0.962692177	1.387608197	-0.345704726	5.80E-20	1.09E-19
gLV (ridge)	gLV (elastic net)	(7.0, 7.3]	1.877331449	1.387608197	0.321839069	8.45E-21	1.69E-20
MDSINE2	MDSINE2 (No Modules)	(7.3, 7.5]	0.945437688	0.910431035	0.043134093	0.003639361	0.003899316
MDSINE2	gLV (ridge)	(7.3, 7.5]	0.945437688	1.913699931	-0.810759277	9.99E-41	4.99E-40
MDSINE2	gLV (elastic net)	(7.3, 7.5]	0.945437688	1.396011795	-0.340575552	1.24E-17	1.90E-17
MDSINE2 (No Modules)	gLV (ridge)	(7.3, 7.5]	0.910431035	1.913699931	-0.879966709	1.31E-50	1.97E-49
MDSINE2 (No Modules)	gLV (elastic net)	(7.3, 7.5]	0.910431035	1.396011795	-0.445583465	1.49E-26	4.24E-26
gLV (ridge)	gLV (elastic net)	(7.3, 7.5]	1.913699931	1.396011795	0.3664111	5.12E-21	1.06E-20
MDSINE2	MDSINE2 (No Modules)	(7.5, 7.8]	0.939195649	0.794499005	0.073095581	7.21E-07	8.32E-07
MDSINE2	gLV (ridge)	(7.5, 7.8]	0.939195649	1.934139778	-0.7334034	2.27E-45	1.36E-44
MDSINE2	gLV (elastic net)	(7.5, 7.8]	0.939195649	1.446937675	-0.351796773	7.44E-22	1.73E-21
MDSINE2 (No Modules)	gLV (ridge)	(7.5, 7.8]	0.794499005	1.934139778	-0.991377675	9.10E-59	2.73E-57
MDSINE2 (No Modules)	gLV (elastic net)	(7.5, 7.8]	0.794499005	1.446937675	-0.502486944	2.84E-35	9.46E-35
gLV (ridge)	gLV (elastic net)	(7.5, 7.8]	1.934139778	1.446937675	0.430344728	1.99E-23	5.19E-23
MDSINE2	MDSINE2 (No Modules)	(7.8, 8.1]	1.00144635	0.757157318	0.117403384	7.57E-12	9.87E-12
MDSINE2	gLV (ridge)	(7.8, 8.1]	1.00144635	2.000071593	-0.771943728	7.05E-47	6.04E-46
MDSINE2	gLV (elastic net)	(7.8, 8.1]	1.00144635	1.491481419	-0.336585517	7.48E-22	1.73E-21
MDSINE2 (No Modules)	gLV (ridge)	(7.8, 8.1]	0.757157318	2.000071593	-1.035062193	5.13E-65	3.08E-63
MDSINE2 (No Modules)	gLV (elastic net)	(7.8, 8.1]	0.757157318	1.491481419	-0.561470522	2.54E-47	2.54E-46
gLV (ridge)	gLV (elastic net)	(7.8, 8.1]	2.000071593	1.491481419	0.316333838	3.35E-20	6.48E-20
MDSINE2	MDSINE2 (No Modules)	(8.1, 8.5]	1.003290794	0.724996903	0.17227368	3.44E-16	5.03E-16
MDSINE2	gLV (ridge)	(8.1, 8.5]	1.003290794	2.02722637	-0.728235337	1.19E-35	4.48E-35
MDSINE2	gLV (elastic net)	(8.1, 8.5]	1.003290794	1.511380116	-0.350164928	2.03E-19	3.58E-19
MDSINE2 (No Modules)	gLV (ridge)	(8.1, 8.5]	0.724996903	2.02722637	-0.940727516	3.80E-58	7.59E-57
MDSINE2 (No Modules)	gLV (elastic net)	(8.1, 8.5]	0.724996903	1.511380116	-0.607433881	8.27E-47	6.20E-46
gLV (ridge)	gLV (elastic net)	(8.1, 8.5]	2.02722637	1.511380116	0.300221173	3.09E-15	4.42E-15
MDSINE2	MDSINE2 (No Modules)	(8.5, 8.9]	1.013185285	0.698898589	0.193045715	1.08E-18	1.86E-18
MDSINE2	gLV (ridge)	(8.5, 8.9]	1.013185285	2.148173906	-0.66885483	2.18E-27	6.55E-27
MDSINE2	gLV (elastic net)	(8.5, 8.9]	1.013185285	1.474180552	-0.304680251	1.56E-11	1.97E-11
MDSINE2 (No Modules)	gLV (ridge)	(8.5, 8.9]	0.698898589	2.148173906	-1.116037448	1.82E-49	2.18E-48
MDSINE2 (No Modules)	gLV (elastic net)	(8.5, 8.9]	0.698898589	1.474180552	-0.597322411	3.67E-38	1.57E-37
gLV (ridge)	gLV (elastic net)	(8.5, 8.9]	2.148173906	1.474180552	0.356356347	5.65E-14	7.89E-14
MDSINE2	MDSINE2 (No Modules)	(8.9, 9.4]	1.174692422	0.61631171	0.30417304	1.36E-21	3.02E-21
MDSINE2	gLV (ridge)	(8.9, 9.4]	1.174692422	2.223195362	-0.56913446	3.45E-18	5.45E-18
MDSINE2	gLV (elastic net)	(8.9, 9.4]	1.174692422	1.480938916	-0.295680665	6.86E-08	8.07E-08
MDSINE2 (No Modules)	gLV (ridge)	(8.9, 9.4]	0.61631171	2.223195362	-1.083858285	1.41E-38	6.52E-38
MDSINE2 (No Modules)	gLV (elastic net)	(8.9, 9.4]	0.61631171	1.480938916	-0.699094417	6.19E-32	1.96E-31
gLV (ridge)	gLV (elastic net)	(8.9, 9.4]	2.223195362	1.480938916	0.309379876	2.23E-10	2.73E-10
MDSINE2	MDSINE2 (No Modules)	(9.4, 11.2]	1.246115583	0.646683991	0.541131034	1.19E-22	2.98E-22
MDSINE2	gLV (ridge)	(9.4, 11.2]	1.246115583	1.941293896	-0.352518839	6.65E-06	7.39E-06
MDSINE2	gLV (elastic net)	(9.4, 11.2]	1.246115583	1.262229709	0.025028323	0.723069831	0.735325252
MDSINE2 (No Modules)	gLV (ridge)	(9.4, 11.2]	0.646683991	1.941293896	-1.00157687	3.28E-21	7.02E-21
MDSINE2 (No Modules)	gLV (elastic net)	(9.4, 11.2]	0.646683991	1.262229709	-0.442464352	3.16E-18	5.12E-18
gLV (ridge)	gLV (elastic net)	(9.4, 11.2]	1.941293896	1.262229709	0.400733778	6.66E-09	7.99E-09

Supplemental Table 5: ***p*****-values associated with Figure 4C.**

Method A	Method B	Median A	Median B	Median A minus B	p	BHCorrectedp
MDSINE2	MDSINE2 (No Modules)	0.957595993	0.8525636	0.081819895	8.80E-09	1.12E-08
MDSINE2	gLV (ridge)	0.957595993	2.406869401	-1.355265884	5.77E-82	4.04E-81
MDSINE2	gLV (elastic net)	0.957595993	1.576258627	-0.516491597	4.17E-57	9.72E-57
MDSINE2	gLV-RA (ridge)	0.957595993	2.851254421	-1.780842	7.15E-80	2.86E-79
MDSINE2	gLV-RA (elastic net)	0.957595993	3.148218679	-2.076211147	1.13E-80	5.26E-80
MDSINE2	cLV	0.957595993	2.440354583	-1.330361648	2.65E-78	9.27E-78
MDSINE2	LRA	0.957595993	5.507975998	-4.174724893	5.03E-71	1.41E-70
MDSINE2 (No Modules)	gLV (ridge)	0.8525636	2.406869401	-1.406978101	1.57E-87	4.41E-86
MDSINE2 (No Modules)	gLV (elastic net)	0.8525636	1.576258627	-0.618572159	1.53E-65	3.90E-65
MDSINE2 (No Modules)	gLV-RA (ridge)	0.8525636	2.851254421	-1.871503505	5.07E-85	7.10E-84
MDSINE2 (No Modules)	gLV-RA (elastic net)	0.8525636	3.148218679	-2.127072799	8.70E-83	8.12E-82
MDSINE2 (No Modules)	cLV	0.8525636	2.440354583	-1.522558093	9.59E-81	5.26E-80
MDSINE2 (No Modules)	LRA	0.8525636	5.507975998	-4.097417397	9.75E-73	3.03E-72
gLV (ridge)	gLV (elastic net)	2.406869401	1.576258627	0.619909899	2.68E-44	4.69E-44
gLV (ridge)	gLV-RA (ridge)	2.406869401	2.851254421	-0.162335126	0.000122049	0.00014239
gLV (ridge)	gLV-RA (elastic net)	2.406869401	3.148218679	-0.386976559	9.43E-07	1.15E-06
gLV (ridge)	cLV	2.406869401	2.440354583	-0.101756525	0.262318692	0.262318692
gLV (ridge)	LRA	2.406869401	5.507975998	-1.750185431	1.03E-25	1.60E-25
gLV (elastic net)	gLV-RA (ridge)	1.576258627	2.851254421	-1.190920049	2.84E-48	5.30E-48
gLV (elastic net)	gLV-RA (elastic net)	1.576258627	3.148218679	-1.407795771	1.53E-48	3.06E-48
gLV (elastic net)	cLV	1.576258627	2.440354583	-0.639269082	3.77E-30	6.21E-30
gLV (elastic net)	LRA	1.576258627	5.507975998	-3.18934168	1.09E-52	2.34E-52
gLV-RA (ridge)	gLV-RA (elastic net)	2.851254421	3.148218679	-0.077225254	0.145522958	0.156717031
gLV-RA (ridge)	cLV	2.851254421	2.440354583	0.06285764	0.035139015	0.039355697
gLV-RA (ridge)	LRA	2.851254421	5.507975998	-1.498749843	7.08E-15	9.91E-15
gLV-RA (elastic net)	cLV	3.148218679	2.440354583	-0.181748402	0.201459803	0.208921277
gLV-RA (elastic net)	LRA	3.148218679	5.507975998	-1.191689473	6.19E-12	8.26E-12
cLV	LRA	2.440354583	5.507975998	-1.539773639	1.15E-22	1.69E-22

Supplemental Table 6: ***p*****-values associated with Figure 4D.**

Method A	Method B	Bin	Median A	Median B	Median A mi	p	BHCorrectedp
MDSINE2	MDSINE2 (No Modules)	(-5.0, -4.1]	0.94459603	0.99236773	0.0237968	0.978467607	0.978467607
MDSINE2	gLV (elastic net)	(-5.0, -4.1]	0.94459603	1.24593858	-0.2500479	4.26E-08	6.38E-08
MDSINE2	gLV (ridge)	(-5.0, -4.1]	0.94459603	2.00171447	-0.8767408	1.51E-42	5.89E-42
MDSINE2	cLV	(-5.0, -4.1]	0.94459603	1.91111905	-0.8134015	1.46E-38	5.04E-38
MDSINE2	gLV-RA (ridge)	(-5.0, -4.1]	0.94459603	2.58848739	-1.5843961	7.66E-54	8.58E-53
MDSINE2	gLV-RA (elastic net)	(-5.0, -4.1]	0.94459603	2.34844374	-1.1388972	1.90E-45	8.58E-45
MDSINE2	LRA	(-5.0, -4.1]	0.94459603	5.45366406	-4.1351656	1.21E-68	3.38E-66
MDSINE2 (No Modules)	gLV (elastic net)	(-5.0, -4.1]	0.99236773	1.24593858	-0.226328	1.67E-05	2.29E-05
MDSINE2 (No Modules)	gLV (ridge)	(-5.0, -4.1]	0.99236773	2.00171447	-0.8512593	6.70E-37	2.18E-36
MDSINE2 (No Modules)	cLV	(-5.0, -4.1]	0.99236773	1.91111905	-0.7680341	5.45E-32	1.56E-31
MDSINE2 (No Modules)	gLV-RA (ridge)	(-5.0, -4.1]	0.99236773	2.58848739	-1.4796684	1.38E-49	8.81E-49
MDSINE2 (No Modules)	gLV-RA (elastic net)	(-5.0, -4.1]	0.99236773	2.34844374	-1.2030162	3.59E-38	1.23E-37
MDSINE2 (No Modules)	LRA	(-5.0, -4.1]	0.99236773	5.45366406	-3.9066726	1.36E-65	1.27E-63
gLV (elastic net)	gLV (ridge)	(-5.0, -4.1]	1.24593858	2.00171447	-0.6071899	2.63E-29	7.02E-29
gLV (elastic net)	cLV	(-5.0, -4.1]	1.24593858	1.91111905	-0.5216739	5.16E-23	1.05E-22
gLV (elastic net)	gLV-RA (ridge)	(-5.0, -4.1]	1.24593858	2.58848739	-1.3086151	7.69E-44	3.12E-43
gLV (elastic net)	gLV-RA (elastic net)	(-5.0, -4.1]	1.24593858	2.34844374	-0.9046222	1.06E-36	3.37E-36
gLV (elastic net)	LRA	(-5.0, -4.1]	1.24593858	5.45366406	-3.3157376	7.26E-66	1.02E-63
gLV (ridge)	cLV	(-5.0, -4.1]	2.00171447	1.91111905	5.89E-05	0.962002917	0.973861235
gLV (ridge)	gLV-RA (ridge)	(-5.0, -4.1]	2.00171447	2.58848739	-0.2433147	1.72E-06	2.43E-06
gLV (ridge)	gLV-RA (elastic net)	(-5.0, -4.1]	2.00171447	2.34844374	-0.2252064	0.005484882	0.007077267
gLV (ridge)	LRA	(-5.0, -4.1]	2.00171447	5.45366406	-2.1085743	1.17E-39	4.08E-39
cLV	gLV-RA (ridge)	(-5.0, -4.1]	1.91111905	2.58848739	-0.1276427	0.000130681	0.000176767
cLV	gLV-RA (elastic net)	(-5.0, -4.1]	1.91111905	2.34844374	0.13699671	0.541780053	0.606793659
cLV	LRA	(-5.0, -4.1]	1.91111905	5.45366406	-1.6086918	1.04E-42	4.12E-42
gLV-RA (ridge)	gLV-RA (elastic net)	(-5.0, -4.1]	2.58848739	2.34844374	0.01312525	0.019855979	0.024491957
gLV-RA (ridge)	LRA	(-5.0, -4.1]	2.58848739	5.45366406	-1.6093364	1.54E-24	3.32E-24
gLV-RA (elastic net)	LRA	(-5.0, -4.1]	2.34844374	5.45366406	-1.6337641	1.62E-29	4.40E-29
MDSINE2	MDSINE2 (No Modules)	(-4.1, -3.8]	0.92966693	0.86173778	0.05728466	0.019322354	0.0239392
MDSINE2	gLV (elastic net)	(-4.1, -3.8]	0.92966693	1.40010285	-0.4424348	1.93E-20	3.79E-20
MDSINE2	gLV (ridge)	(-4.1, -3.8]	0.92966693	2.02297545	-0.9618363	1.57E-47	8.64E-47
MDSINE2	cLV	(-4.1, -3.8]	0.92966693	1.91090719	-1.0172519	2.69E-45	1.18E-44
MDSINE2	gLV-RA (ridge)	(-4.1, -3.8]	0.92966693	2.35164921	-1.3881242	1.50E-48	9.10E-48
MDSINE2	gLV-RA (elastic net)	(-4.1, -3.8]	0.92966693	2.21770226	-1.1992357	6.73E-44	2.77E-43
MDSINE2	LRA	(-4.1, -3.8]	0.92966693	5.99164002	-4.6862961	3.82E-63	2.14E-61
MDSINE2 (No Modules)	gLV (elastic net)	(-4.1, -3.8]	0.86173778	1.40010285	-0.4688214	3.06E-24	6.49E-24
MDSINE2 (No Modules)	gLV (ridge)	(-4.1, -3.8]	0.86173778	2.02297545	-0.9942635	1.35E-49	8.81E-49
MDSINE2 (No Modules)	cLV	(-4.1, -3.8]	0.86173778	1.91090719	-1.0133887	5.33E-47	2.81E-46
MDSINE2 (No Modules)	gLV-RA (ridge)	(-4.1, -3.8]	0.86173778	2.35164921	-1.3579718	4.45E-51	3.89E-50
MDSINE2 (No Modules)	gLV-RA (elastic net)	(-4.1, -3.8]	0.86173778	2.21770226	-1.1549585	6.38E-46	3.03E-45
MDSINE2 (No Modules)	LRA	(-4.1, -3.8]	0.86173778	5.99164002	-4.7194393	4.18E-64	2.92E-62
gLV (elastic net)	gLV (ridge)	(-4.1, -3.8]	1.40010285	2.02297545	-0.4675995	3.28E-25	7.47E-25
gLV (elastic net)	cLV	(-4.1, -3.8]	1.40010285	1.91090719	-0.3821756	9.43E-19	1.76E-18
gLV (elastic net)	gLV-RA (ridge)	(-4.1, -3.8]	1.40010285	2.35164921	-0.8922736	7.47E-29	1.88E-28
gLV (elastic net)	gLV-RA (elastic net)	(-4.1, -3.8]	1.40010285	2.21770226	-0.8007604	3.42E-24	7.20E-24
gLV (elastic net)	LRA	(-4.1, -3.8]	1.40010285	5.99164002	-3.7532198	2.28E-59	9.12E-58
gLV (ridge)	cLV	(-4.1, -3.8]	2.02297545	1.91090719	-0.0164771	0.791166563	0.827552382
gLV (ridge)	gLV-RA (ridge)	(-4.1, -3.8]	2.02297545	2.35164921	-0.1864424	0.00654917	0.008373367
gLV (ridge)	gLV-RA (elastic net)	(-4.1, -3.8]	2.02297545	2.21770226	-0.0081341	0.032991586	0.04016367
gLV (ridge)	LRA	(-4.1, -3.8]	2.02297545	5.99164002	-2.3156146	7.11E-41	2.55E-40
cLV	gLV-RA (ridge)	(-4.1, -3.8]	1.91090719	2.35164921	-0.0322616	0.033492531	0.040597007
cLV	gLV-RA (elastic net)	(-4.1, -3.8]	1.91090719	2.21770226	0.19960343	0.816351045	0.849733429
cLV	LRA	(-4.1, -3.8]	1.91090719	5.99164002	-2.1799576	8.88E-40	3.15E-39
gLV-RA (ridge)	gLV-RA (elastic net)	(-4.1, -3.8]	2.35164921	2.21770226	-0.009254	0.792085851	0.827552382
gLV-RA (ridge)	LRA	(-4.1, -3.8]	2.35164921	5.99164002	-2.0523689	2.89E-29	7.63E-29
gLV-RA (elastic net)	LRA	(-4.1, -3.8]	2.21770226	5.99164002	-1.9266801	1.13E-28	2.82E-28
MDSINE2	MDSINE2 (No Modules)	(-3.8, -3.6]	0.9755943	0.86109682	0.09568095	9.57E-06	1.32E-05
MDSINE2	gLV (elastic net)	(-3.8, -3.6]	0.9755943	1.44626047	-0.4267024	2.14E-21	4.25E-21
MDSINE2	gLV (ridge)	(-3.8, -3.6]	0.9755943	2.17766618	-1.0601139	1.35E-50	1.05E-49
MDSINE2	cLV	(-3.8, -3.6]	0.9755943	2.02272469	-1.1016024	4.26E-46	2.09E-45
MDSINE2	gLV-RA (ridge)	(-3.8, -3.6]	0.9755943	2.43921657	-1.5126212	6.41E-50	4.27E-49
MDSINE2	gLV-RA (elastic net)	(-3.8, -3.6]	0.9755943	2.42295634	-1.3573408	1.73E-45	7.95E-45
MDSINE2	LRA	(-3.8, -3.6]	0.9755943	6.26519218	-4.860366	7.93E-59	2.47E-57
MDSINE2 (No Modules)	gLV (elastic net)	(-3.8, -3.6]	0.86109682	1.44626047	-0.4858195	7.04E-29	1.79E-28
MDSINE2 (No Modules)	gLV (ridge)	(-3.8, -3.6]	0.86109682	2.17766618	-1.159643	7.95E-57	1.59E-55
MDSINE2 (No Modules)	cLV	(-3.8, -3.6]	0.86109682	2.02272469	-1.2148116	3.78E-51	3.41E-50
MDSINE2 (No Modules)	gLV-RA (ridge)	(-3.8, -3.6]	0.86109682	2.43921657	-1.5234142	3.57E-54	4.17E-53
MDSINE2 (No Modules)	gLV-RA (elastic net)	(-3.8, -3.6]	0.86109682	2.42295634	-1.4343713	3.43E-50	2.34E-49
MDSINE2 (No Modules)	LRA	(-3.8, -3.6]	0.86109682	6.26519218	-4.9842276	8.65E-60	4.04E-58
gLV (elastic net)	gLV (ridge)	(-3.8, -3.6]	1.44626047	2.17766618	-0.5161453	5.11E-29	1.31E-28
gLV (elastic net)	cLV	(-3.8, -3.6]	1.44626047	2.02272469	-0.5695728	3.43E-20	6.62E-20
gLV (elastic net)	gLV-RA (ridge)	(-3.8, -3.6]	1.44626047	2.43921657	-0.9405123	4.61E-28	1.13E-27
gLV (elastic net)	gLV-RA (elastic net)	(-3.8, -3.6]	1.44626047	2.42295634	-0.9262298	2.11E-25	4.83E-25
gLV (elastic net)	LRA	(-3.8, -3.6]	1.44626047	6.26519218	-3.8224458	3.62E-55	5.34E-54
gLV (ridge)	cLV	(-3.8, -3.6]	2.17766618	2.02272469	0.00020912	0.574756624	0.638618471
gLV (ridge)	gLV-RA (ridge)	(-3.8, -3.6]	2.17766618	2.43921657	-0.1165475	0.004532125	0.005874977
gLV (ridge)	gLV-RA (elastic net)	(-3.8, -3.6]	2.17766618	2.42295634	-0.182878	0.006130868	0.007874509
gLV (ridge)	LRA	(-3.8, -3.6]	2.17766618	6.26519218	-2.5156696	2.32E-35	7.13E-35
cLV	gLV-RA (ridge)	(-3.8, -3.6]	2.02272469	2.43921657	-0.0813865	0.095193355	0.112464723
cLV	gLV-RA (elastic net)	(-3.8, -3.6]	2.02272469	2.42295634	0.06841431	0.971254608	0.974735807
cLV	LRA	(-3.8, -3.6]	2.02272469	6.26519218	-1.9617988	1.62E-33	4.79E-33
gLV-RA (ridge)	gLV-RA (elastic net)	(-3.8, -3.6]	2.43921657	2.42295634	0	0.874865829	0.900597176
gLV-RA (ridge)	LRA	(-3.8, -3.6]	2.43921657	6.26519218	-2.0103196	5.14E-25	1.15E-24
gLV-RA (elastic net)	LRA	(-3.8, -3.6]	2.42295634	6.26519218	-1.8472867	1.75E-23	3.64E-23
MDSINE2	MDSINE2 (No Modules)	(-3.6, -3.3]	0.95525444	0.84613197	0.11865964	1.07E-08	1.62E-08
MDSINE2	gLV (elastic net)	(-3.6, -3.3]	0.95525444	1.49104041	-0.4466475	9.12E-25	1.98E-24
MDSINE2	gLV (ridge)	(-3.6, -3.3]	0.95525444	2.14231457	-0.9692165	2.33E-50	1.67E-49
MDSINE2	cLV	(-3.6, -3.3]	0.95525444	2.23460665	-1.230958	5.32E-51	4.38E-50
MDSINE2	gLV-RA (ridge)	(-3.6, -3.3]	0.95525444	2.18125974	-1.2021182	4.54E-46	2.19E-45
MDSINE2	gLV-RA (elastic net)	(-3.6, -3.3]	0.95525444	2.35000225	-1.3597026	3.98E-47	2.14E-46
MDSINE2	LRA	(-3.6, -3.3]	0.95525444	6.47468662	-4.9175061	1.66E-54	2.21E-53
MDSINE2 (No Modules)	gLV (elastic net)	(-3.6, -3.3]	0.84613197	1.49104041	-0.5282395	3.29E-31	9.32E-31
MDSINE2 (No Modules)	gLV (ridge)	(-3.6, -3.3]	0.84613197	2.14231457	-1.2409025	2.69E-54	3.28E-53
MDSINE2 (No Modules)	cLV	(-3.6, -3.3]	0.84613197	2.23460665	-1.3677979	1.11E-56	2.08E-55
MDSINE2 (No Modules)	gLV-RA (ridge)	(-3.6, -3.3]	0.84613197	2.18125974	-1.3317807	1.51E-50	1.14E-49
MDSINE2 (No Modules)	gLV-RA (elastic net)	(-3.6, -3.3]	0.84613197	2.35000225	-1.5026024	2.69E-53	2.89E-52
MDSINE2 (No Modules)	LRA	(-3.6, -3.3]	0.84613197	6.47468662	-5.0049255	7.19E-57	1.55E-55
gLV (elastic net)	gLV (ridge)	(-3.6, -3.3]	1.49104041	2.14231457	-0.4528403	7.89E-25	1.74E-24
gLV (elastic net)	cLV	(-3.6, -3.3]	1.49104041	2.23460665	-0.6186632	1.18E-22	2.38E-22
gLV (elastic net)	gLV-RA (ridge)	(-3.6, -3.3]	1.49104041	2.18125974	-0.7067803	1.98E-20	3.85E-20
gLV (elastic net)	gLV-RA (elastic net)	(-3.6, -3.3]	1.49104041	2.35000225	-0.8431636	5.54E-25	1.23E-24
gLV (elastic net)	LRA	(-3.6, -3.3]	1.49104041	6.47468662	-3.9662943	3.03E-50	2.12E-49
gLV (ridge)	cLV	(-3.6, -3.3]	2.14231457	2.23460665	-0.127376	0.059826842	0.071894917
gLV (ridge)	gLV-RA (ridge)	(-3.6, -3.3]	2.14231457	2.18125974	-0.091215	0.080346703	0.096141354
gLV (ridge)	gLV-RA (elastic net)	(-3.6, -3.3]	2.14231457	2.35000225	-0.107657	0.002343437	0.003080574
gLV (ridge)	LRA	(-3.6, -3.3]	2.14231457	6.47468662	-2.4408153	3.41E-34	1.03E-33
cLV	gLV-RA (ridge)	(-3.6, -3.3]	2.23460665	2.18125974	0.11327648	0.593647044	0.651847735
cLV	gLV-RA (elastic net)	(-3.6, -3.3]	2.23460665	2.35000225	0.15009811	0.586665067	0.649273592
cLV	LRA	(-3.6, -3.3]	2.23460665	6.47468662	-1.9531507	3.12E-27	7.54E-27
gLV-RA (ridge)	gLV-RA (elastic net)	(-3.6, -3.3]	2.18125974	2.35000225	-0.0415906	0.03167782	0.038732705
gLV-RA (ridge)	LRA	(-3.6, -3.3]	2.18125974	6.47468662	-1.9741048	4.90E-24	1.02E-23
gLV-RA (elastic net)	LRA	(-3.6, -3.3]	2.35000225	6.47468662	-1.8860466	8.20E-20	1.55E-19
MDSINE2	MDSINE2 (No Modules)	(-3.3, -3.1]	0.92673266	0.78363739	0.13061438	1.61E-09	2.51E-09
MDSINE2	gLV (elastic net)	(-3.3, -3.1]	0.92673266	1.53682588	-0.4854472	4.18E-29	1.09E-28
MDSINE2	gLV (ridge)	(-3.3, -3.1]	0.92673266	2.21228731	-1.1003815	2.02E-52	2.10E-51
MDSINE2	cLV	(-3.3, -3.1]	0.92673266	2.27416938	-1.1950289	7.11E-52	6.86E-51
MDSINE2	gLV-RA (ridge)	(-3.3, -3.1]	0.92673266	2.16061722	-1.2202299	2.57E-45	1.14E-44
MDSINE2	gLV-RA (elastic net)	(-3.3, -3.1]	0.92673266	2.3786758	-1.3970848	3.23E-48	1.88E-47
MDSINE2	LRA	(-3.3, -3.1]	0.92673266	6.67859418	-4.7642957	3.95E-56	6.51E-55
MDSINE2 (No Modules)	gLV (elastic net)	(-3.3, -3.1]	0.78363739	1.53682588	-0.634692	4.16E-41	1.51E-40
MDSINE2 (No Modules)	gLV (ridge)	(-3.3, -3.1]	0.78363739	2.21228731	-1.3252312	2.62E-59	9.15E-58
MDSINE2 (No Modules)	cLV	(-3.3, -3.1]	0.78363739	2.27416938	-1.4379795	1.13E-58	3.15E-57
MDSINE2 (No Modules)	gLV-RA (ridge)	(-3.3, -3.1]	0.78363739	2.16061722	-1.2908527	2.31E-52	2.31E-51
MDSINE2 (No Modules)	gLV-RA (elastic net)	(-3.3, -3.1]	0.78363739	2.3786758	-1.5467639	2.58E-54	3.28E-53
MDSINE2 (No Modules)	LRA	(-3.3, -3.1]	0.78363739	6.67859418	-5.068636	6.86E-58	1.60E-56
gLV (elastic net)	gLV (ridge)	(-3.3, -3.1]	1.53682588	2.21228731	-0.4924288	2.07E-25	4.80E-25
gLV (elastic net)	cLV	(-3.3, -3.1]	1.53682588	2.27416938	-0.5374571	7.81E-20	1.49E-19
gLV (elastic net)	gLV-RA (ridge)	(-3.3, -3.1]	1.53682588	2.16061722	-0.596228	8.99E-18	1.61E-17
gLV (elastic net)	gLV-RA (elastic net)	(-3.3, -3.1]	1.53682588	2.3786758	-0.8731246	1.95E-24	4.17E-24
gLV (elastic net)	LRA	(-3.3, -3.1]	1.53682588	6.67859418	-3.9276269	4.96E-49	3.09E-48
gLV (ridge)	cLV	(-3.3, -3.1]	2.21228731	2.27416938	-0.0889475	0.227605554	0.262261544
gLV (ridge)	gLV-RA (ridge)	(-3.3, -3.1]	2.21228731	2.16061722	-0.0321991	0.293172327	0.33642726
gLV (ridge)	gLV-RA (elastic net)	(-3.3, -3.1]	2.21228731	2.3786758	-0.1339714	0.002786469	0.003645847
gLV (ridge)	LRA	(-3.3, -3.1]	2.21228731	6.67859418	-2.4152516	4.64E-30	1.30E-29
cLV	gLV-RA (ridge)	(-3.3, -3.1]	2.27416938	2.16061722	0.11456567	0.624666297	0.683228763
cLV	gLV-RA (elastic net)	(-3.3, -3.1]	2.27416938	2.3786758	0.1333597	0.935247631	0.955727506
cLV	LRA	(-3.3, -3.1]	2.27416938	6.67859418	-1.8238507	1.73E-25	4.04E-25
gLV-RA (ridge)	gLV-RA (elastic net)	(-3.3, -3.1]	2.16061722	2.3786758	-0.052425	0.007905027	0.009970305
gLV-RA (ridge)	LRA	(-3.3, -3.1]	2.16061722	6.67859418	-1.9289431	1.55E-22	3.10E-22
gLV-RA (elastic net)	LRA	(-3.3, -3.1]	2.3786758	6.67859418	-1.5684415	4.00E-18	7.26E-18

Method A	Method B	Bin	Median A	Median B	Median A minus B	p	BHCorrectedp
MDSINE2	MDSINE2 (No Mod	(-3.1, -2.8]	0.911933233	0.822978419	0.091111286	6.65E-06	9.26E-06
MDSINE2	gLV (elastic net)	(-3.1, -2.8]	0.911933233	1.634645681	-0.548991555	1.69E-35	5.25E-35
MDSINE2	gLV (ridge)	(-3.1, -2.8]	0.911933233	2.47017968	-1.336561608	9.14E-55	1.28E-53
MDSINE2	cLV	(-3.1, -2.8]	0.911933233	2.435938485	-1.423035799	8.08E-52	7.54E-51
MDSINE2	gLV-RA (ridge)	(-3.1, -2.8]	0.911933233	2.378492568	-1.392743936	2.80E-43	1.12E-42
MDSINE2	gLV-RA (elastic net	(-3.1, -2.8]	0.911933233	2.469182719	-1.529733101	2.16E-46	1.08E-45
MDSINE2	LRA	(-3.1, -2.8]	0.911933233	5.074711399	-4.191479155	1.32E-46	6.83E-46
MDSINE2 (No Mod	gLV (elastic net)	(-3.1, -2.8]	0.822978419	1.634645681	-0.643939096	8.50E-42	3.17E-41
MDSINE2 (No Mod	gLV (ridge)	(-3.1, -2.8]	0.822978419	2.47017968	-1.393200977	2.15E-58	5.47E-57
MDSINE2 (No Mod	cLV	(-3.1, -2.8]	0.822978419	2.435938485	-1.548228191	1.23E-56	2.16E-55
MDSINE2 (No Mod	gLV-RA (ridge)	(-3.1, -2.8]	0.822978419	2.378492568	-1.52871541	4.31E-48	2.46E-47
MDSINE2 (No Mod	gLV-RA (elastic net	(-3.1, -2.8]	0.822978419	2.469182719	-1.673172404	5.09E-51	4.32E-50
MDSINE2 (No Mod	LRA	(-3.1, -2.8]	0.822978419	5.074711399	-4.196332185	1.65E-46	8.40E-46
gLV (elastic net)	gLV (ridge)	(-3.1, -2.8]	1.634645681	2.47017968	-0.565572668	7.21E-27	1.73E-26
gLV (elastic net)	cLV	(-3.1, -2.8]	1.634645681	2.435938485	-0.545760338	1.16E-18	2.15E-18
gLV (elastic net)	gLV-RA (ridge)	(-3.1, -2.8]	1.634645681	2.378492568	-0.661409678	4.59E-18	8.29E-18
gLV (elastic net)	gLV-RA (elastic net	(-3.1, -2.8]	1.634645681	2.469182719	-0.900624334	3.88E-23	7.98E-23
gLV (elastic net)	LRA	(-3.1, -2.8]	1.634645681	5.074711399	-3.431203887	1.67E-35	5.25E-35
gLV (ridge)	cLV	(-3.1, -2.8]	2.47017968	2.435938485	-0.097262539	0.412510009	0.47144001
gLV (ridge)	gLV-RA (ridge)	(-3.1, -2.8]	2.47017968	2.378492568	0.163134963	0.963427007	0.973861235
gLV (ridge)	gLV-RA (elastic net	(-3.1, -2.8]	2.47017968	2.469182719	-0.136715924	0.007211046	0.009136167
gLV (ridge)	LRA	(-3.1, -2.8]	2.47017968	5.074711399	-1.794532413	2.09E-16	3.68E-16
cLV	gLV-RA (ridge)	(-3.1, -2.8]	2.435938485	2.378492568	0.146080686	0.527287946	0.595325101
cLV	gLV-RA (elastic net	(-3.1, -2.8]	2.435938485	2.469182719	0.116926648	0.675824033	0.730620577
cLV	LRA	(-3.1, -2.8]	2.435938485	5.074711399	-1.043234329	4.40E-13	7.33E-13
gLV-RA (ridge)	gLV-RA (elastic net	(-3.1, -2.8]	2.378492568	2.469182719	-0.03668981	0.01389709	0.017294157
gLV-RA (ridge)	LRA	(-3.1, -2.8]	2.378492568	5.074711399	-1.216014637	9.87E-12	1.58E-11
gLV-RA (elastic net	LRA	(-3.1, -2.8]	2.469182719	5.074711399	-0.773018954	1.18E-07	1.74E-07
MDSINE2	MDSINE2 (No Mod	(-2.8, -2.4]	0.882298995	0.711656949	0.085146212	8.12E-07	1.17E-06
MDSINE2	gLV (elastic net)	(-2.8, -2.4]	0.882298995	1.719578774	-0.587535861	8.53E-35	2.60E-34
MDSINE2	gLV (ridge)	(-2.8, -2.4]	0.882298995	2.601909765	-1.354419052	1.56E-50	1.15E-49
MDSINE2	cLV	(-2.8, -2.4]	0.882298995	2.467755277	-1.283451261	8.97E-46	4.18E-45
MDSINE2	gLV-RA (ridge)	(-2.8, -2.4]	0.882298995	2.527744432	-1.372749184	6.79E-37	2.19E-36
MDSINE2	gLV-RA (elastic net	(-2.8, -2.4]	0.882298995	2.573303648	-1.381802294	5.44E-42	2.09E-41
MDSINE2	LRA	(-2.8, -2.4]	0.882298995	2.309494703	-1.534805172	7.61E-30	2.09E-29
MDSINE2 (No Mod	gLV (elastic net)	(-2.8, -2.4]	0.711656949	1.719578774	-0.684161275	2.43E-44	1.05E-43
MDSINE2 (No Mod	gLV (ridge)	(-2.8, -2.4]	0.711656949	2.601909765	-1.583756791	5.98E-56	9.30E-55
MDSINE2 (No Mod	cLV	(-2.8, -2.4]	0.711656949	2.467755277	-1.500757803	1.29E-50	1.03E-49
MDSINE2 (No Mod	gLV-RA (ridge)	(-2.8, -2.4]	0.711656949	2.527744432	-1.490101819	5.13E-44	2.18E-43
MDSINE2 (No Mod	gLV-RA (elastic net	(-2.8, -2.4]	0.711656949	2.573303648	-1.709441094	8.56E-48	4.79E-47
MDSINE2 (No Mod	LRA	(-2.8, -2.4]	0.711656949	2.309494703	-1.548100455	6.51E-33	1.88E-32
gLV (elastic net)	gLV (ridge)	(-2.8, -2.4]	1.719578774	2.601909765	-0.57034583	7.65E-23	1.55E-22
gLV (elastic net)	cLV	(-2.8, -2.4]	1.719578774	2.467755277	-0.419310621	1.14E-12	1.88E-12
gLV (elastic net)	gLV-RA (ridge)	(-2.8, -2.4]	1.719578774	2.527744432	-0.426708878	3.38E-12	5.53E-12
gLV (elastic net)	gLV-RA (elastic net	(-2.8, -2.4]	1.719578774	2.573303648	-0.503668394	5.93E-14	1.01E-13
gLV (elastic net)	LRA	(-2.8, -2.4]	1.719578774	2.309494703	-0.74268045	9.69E-18	1.73E-17
gLV (ridge)	cLV	(-2.8, -2.4]	2.601909765	2.467755277	-0.039579603	0.628916979	0.685201378
gLV (ridge)	gLV-RA (ridge)	(-2.8, -2.4]	2.601909765	2.527744432	0.290158802	0.774740094	0.818593307
gLV (ridge)	gLV-RA (elastic net	(-2.8, -2.4]	2.601909765	2.573303648	0.088645148	0.141938725	0.166288046
gLV (ridge)	LRA	(-2.8, -2.4]	2.601909765	2.309494703	-0.520727955	1.48E-06	2.10E-06
cLV	gLV-RA (ridge)	(-2.8, -2.4]	2.467755277	2.527744432	0.146086717	0.692315809	0.742714278
cLV	gLV-RA (elastic net	(-2.8, -2.4]	2.467755277	2.573303648	0.12738429	0.540775918	0.606793659
cLV	LRA	(-2.8, -2.4]	2.467755277	2.309494703	-0.281824908	7.23E-05	9.83E-05
gLV-RA (ridge)	gLV-RA (elastic net	(-2.8, -2.4]	2.527744432	2.573303648	-0.086841481	0.093167932	0.110610858
gLV-RA (ridge)	LRA	(-2.8, -2.4]	2.527744432	2.309494703	-0.198683336	0.000387442	0.000516589
gLV-RA (elastic net	LRA	(-2.8, -2.4]	2.573303648	2.309494703	-0.042970182	0.006787473	0.008638602
MDSINE2	MDSINE2 (No Mod	(-2.4, -2.0]	0.886348366	0.715970076	0.100969246	1.40E-07	2.06E-07
MDSINE2	gLV (elastic net)	(-2.4, -2.0]	0.886348366	1.676703737	-0.568497086	4.42E-25	9.97E-25
MDSINE2	gLV (ridge)	(-2.4, -2.0]	0.886348366	2.735839277	-1.22795831	6.51E-42	2.46E-41
MDSINE2	cLV	(-2.4, -2.0]	0.886348366	2.578636709	-1.078267362	4.11E-38	1.39E-37
MDSINE2	gLV-RA (ridge)	(-2.4, -2.0]	0.886348366	2.446743344	-1.367968735	1.71E-28	4.23E-28
MDSINE2	gLV-RA (elastic net	(-2.4, -2.0]	0.886348366	2.658622707	-1.25286927	4.02E-33	1.17E-32
MDSINE2	LRA	(-2.4, -2.0]	0.886348366	1.065991327	-0.276484251	3.05E-14	5.20E-14
MDSINE2 (No Mod	gLV (elastic net)	(-2.4, -2.0]	0.715970076	1.676703737	-0.747876171	3.15E-37	1.04E-36
MDSINE2 (No Mod	gLV (ridge)	(-2.4, -2.0]	0.715970076	2.735839277	-1.481276946	1.53E-48	9.10E-48
MDSINE2 (No Mod	cLV	(-2.4, -2.0]	0.715970076	2.578636709	-1.277853365	5.26E-44	2.20E-43
MDSINE2 (No Mod	gLV-RA (ridge)	(-2.4, -2.0]	0.715970076	2.446743344	-1.458692946	1.49E-37	4.97E-37
MDSINE2 (No Mod	gLV-RA (elastic net	(-2.4, -2.0]	0.715970076	2.658622707	-1.409010366	1.53E-41	5.64E-41
MDSINE2 (No Mod	LRA	(-2.4, -2.0]	0.715970076	1.065991327	-0.377015092	1.74E-18	3.18E-18
gLV (elastic net)	gLV (ridge)	(-2.4, -2.0]	1.676703737	2.735839277	-0.532760224	4.04E-20	7.76E-20
gLV (elastic net)	cLV	(-2.4, -2.0]	1.676703737	2.578636709	-0.427524569	5.22E-11	8.25E-11
gLV (elastic net)	gLV-RA (ridge)	(-2.4, -2.0]	1.676703737	2.446743344	-0.524587662	3.86E-11	6.14E-11
gLV (elastic net)	gLV-RA (elastic net	(-2.4, -2.0]	1.676703737	2.658622707	-0.393105829	1.31E-10	2.05E-10
gLV (elastic net)	LRA	(-2.4, -2.0]	1.676703737	1.065991327	-0.121324766	2.60E-06	3.66E-06
gLV (ridge)	cLV	(-2.4, -2.0]	2.735839277	2.578636709	-0.000403694	0.86088334	0.889473562
gLV (ridge)	gLV-RA (ridge)	(-2.4, -2.0]	2.735839277	2.446743344	0.285805062	0.519148278	0.58850817
gLV (ridge)	gLV-RA (elastic net	(-2.4, -2.0]	2.735839277	2.658622707	0.062996193	0.436397838	0.496712987
gLV (ridge)	LRA	(-2.4, -2.0]	2.735839277	1.065991327	0.158220782	0.194060426	0.225464396
cLV	gLV-RA (ridge)	(-2.4, -2.0]	2.578636709	2.446743344	0.145598894	0.750529608	0.799042929
cLV	gLV-RA (elastic net	(-2.4, -2.0]	2.578636709	2.658622707	0.26015978	0.093229151	0.110610858
cLV	LRA	(-2.4, -2.0]	2.578636709	1.065991327	0.144387844	0.828531781	0.859218143
gLV-RA (ridge)	gLV-RA (elastic net	(-2.4, -2.0]	2.446743344	2.658622707	-0.001123955	0.646463793	0.701588613
gLV-RA (ridge)	LRA	(-2.4, -2.0]	2.446743344	1.065991327	0.056742252	0.734820112	0.785303936
gLV-RA (elastic net	LRA	(-2.4, -2.0]	2.658622707	1.065991327	0.158514271	0.968440962	0.974735807
MDSINE2	MDSINE2 (No Mod	(-2.0, -1.5]	0.753961476	0.589035744	0.133350332	4.99E-07	7.24E-07
MDSINE2	gLV (elastic net)	(-2.0, -1.5]	0.753961476	1.605479989	-0.571407813	1.62E-18	2.99E-18
MDSINE2	gLV (ridge)	(-2.0, -1.5]	0.753961476	2.708679911	-1.245274266	1.03E-27	2.51E-27
MDSINE2	cLV	(-2.0, -1.5]	0.753961476	2.565992239	-1.189416857	2.83E-26	6.72E-26
MDSINE2	gLV-RA (ridge)	(-2.0, -1.5]	0.753961476	2.919962861	-1.671227498	1.45E-20	2.87E-20
MDSINE2	gLV-RA (elastic net	(-2.0, -1.5]	0.753961476	2.78177736	-1.243287983	8.15E-25	1.78E-24
MDSINE2	LRA	(-2.0, -1.5]	0.753961476	0.676270228	-0.034722872	0.001843005	0.002434157
MDSINE2 (No Mod	gLV (elastic net)	(-2.0, -1.5]	0.589035744	1.605479989	-0.806873453	4.70E-30	1.30E-29
MDSINE2 (No Mod	gLV (ridge)	(-2.0, -1.5]	0.589035744	2.708679911	-1.389318568	1.45E-33	4.31E-33
MDSINE2 (No Mod	cLV	(-2.0, -1.5]	0.589035744	2.565992239	-1.438803372	1.92E-29	5.16E-29
MDSINE2 (No Mod	gLV-RA (ridge)	(-2.0, -1.5]	0.589035744	2.919962861	-1.908821212	8.08E-26	1.90E-25
MDSINE2 (No Mod	gLV-RA (elastic net	(-2.0, -1.5]	0.589035744	2.78177736	-1.834887032	5.01E-29	1.30E-28
MDSINE2 (No Mod	LRA	(-2.0, -1.5]	0.589035744	0.676270228	-0.132334732	6.21E-06	8.70E-06
gLV (elastic net)	gLV (ridge)	(-2.0, -1.5]	1.605479989	2.708679911	-0.479358766	3.87E-13	6.49E-13
gLV (elastic net)	cLV	(-2.0, -1.5]	1.605479989	2.565992239	-0.330877186	6.71E-07	9.68E-07
gLV (elastic net)	gLV-RA (ridge)	(-2.0, -1.5]	1.605479989	2.919962861	-0.729984239	3.04E-09	4.68E-09
gLV (elastic net)	gLV-RA (elastic net	(-2.0, -1.5]	1.605479989	2.78177736	-0.607657628	1.73E-09	2.68E-09
gLV (elastic net)	LRA	(-2.0, -1.5]	1.605479989	0.676270228	0.35523924	0.589105449	0.649407582
gLV (ridge)	cLV	(-2.0, -1.5]	2.708679911	2.565992239	0.04537814	0.939384457	0.956464175
gLV (ridge)	gLV-RA (ridge)	(-2.0, -1.5]	2.708679911	2.919962861	-0.000120426	0.141716299	0.166288046
gLV (ridge)	gLV-RA (elastic net	(-2.0, -1.5]	2.708679911	2.78177736	0.055129199	0.145389809	0.169621444
gLV (ridge)	LRA	(-2.0, -1.5]	2.708679911	0.676270228	0.825372183	0.000287171	0.000386576
cLV	gLV-RA (ridge)	(-2.0, -1.5]	2.565992239	2.919962861	-0.004022579	0.226713309	0.262261544
cLV	gLV-RA (elastic net	(-2.0, -1.5]	2.565992239	2.78177736	0.004216697	0.687254682	0.740120427
cLV	LRA	(-2.0, -1.5]	2.565992239	0.676270228	0.709844207	0.000373134	0.000499893
gLV-RA (ridge)	gLV-RA (elastic net	(-2.0, -1.5]	2.919962861	2.78177736	-0.000916766	0.783555503	0.824795266
gLV-RA (ridge)	LRA	(-2.0, -1.5]	2.919962861	0.676270228	0.90253267	8.80E-06	1.22E-05
gLV-RA (elastic net	LRA	(-2.0, -1.5]	2.78177736	0.676270228	0.84468371	1.36E-05	1.86E-05
MDSINE2	MDSINE2 (No Mod	(-1.5, -0.2]	0.732823461	0.411349293	0.162744484	1.67E-07	2.45E-07
MDSINE2	gLV (elastic net)	(-1.5, -0.2]	0.732823461	1.297710687	-0.446614942	1.15E-06	1.64E-06
MDSINE2	gLV (ridge)	(-1.5, -0.2]	0.732823461	1.977810175	-0.989446835	5.94E-15	1.03E-14
MDSINE2	cLV	(-1.5, -0.2]	0.732823461	1.948268377	-0.958464212	3.64E-13	6.15E-13
MDSINE2	gLV-RA (ridge)	(-1.5, -0.2]	0.732823461	2.818718931	-1.994814952	7.29E-12	1.18E-11
MDSINE2	gLV-RA (elastic net	(-1.5, -0.2]	0.732823461	2.766408636	-1.376379526	8.02E-13	1.33E-12
MDSINE2	LRA	(-1.5, -0.2]	0.732823461	0.437488491	0.141495281	0.004254334	0.005540528
MDSINE2 (No Mod	gLV (elastic net)	(-1.5, -0.2]	0.411349293	1.297710687	-0.624724623	3.76E-17	6.66E-17
MDSINE2 (No Mod	gLV (ridge)	(-1.5, -0.2]	0.411349293	1.977810175	-1.175334029	4.24E-19	7.97E-19
MDSINE2 (No Mod	cLV	(-1.5, -0.2]	0.411349293	1.948268377	-1.208513096	3.48E-15	6.05E-15
MDSINE2 (No Mod	gLV-RA (ridge)	(-1.5, -0.2]	0.411349293	2.818718931	-2.324406973	7.99E-15	1.37E-14
MDSINE2 (No Mod	gLV-RA (elastic net	(-1.5, -0.2]	0.411349293	2.766408636	-1.913254651	2.46E-16	4.30E-16
MDSINE2 (No Mod	LRA	(-1.5, -0.2]	0.411349293	0.437488491	-0.003113616	0.764613805	0.810954036
gLV (elastic net)	gLV (ridge)	(-1.5, -0.2]	1.297710687	1.977810175	-0.635388748	6.04E-09	9.24E-09
gLV (elastic net)	cLV	(-1.5, -0.2]	1.297710687	1.948268377	-0.374680641	0.000856112	0.001136072
gLV (elastic net)	gLV-RA (ridge)	(-1.5, -0.2]	1.297710687	2.818718931	-1.159980312	5.27E-08	7.84E-08
gLV (elastic net)	gLV-RA (elastic net	(-1.5, -0.2]	1.297710687	2.766408636	-1.052361127	1.16E-08	1.76E-08
gLV (elastic net)	LRA	(-1.5, -0.2]	1.297710687	0.437488491	0.570903958	3.73E-07	5.44E-07
gLV (ridge)	cLV	(-1.5, -0.2]	1.977810175	1.948268377	0.000134604	0.913392297	0.936812612
gLV (ridge)	gLV-RA (ridge)	(-1.5, -0.2]	1.977810175	2.818718931	-0.13029418	0.01996836	0.024522548
gLV (ridge)	gLV-RA (elastic net	(-1.5, -0.2]	1.977810175	2.766408636	-0.18331678	0.012843888	0.01605486
gLV (ridge)	LRA	(-1.5, -0.2]	1.977810175	0.437488491	1.060995506	8.17E-11	1.29E-10
cLV	gLV-RA (ridge)	(-1.5, -0.2]	1.948268377	2.818718931	-0.301314798	0.012180655	0.015294096
cLV	gLV-RA (elastic net	(-1.5, -0.2]	1.948268377	2.766408636	-0.00013461	0.04499642	0.054306024
cLV	LRA	(-1.5, -0.2]	1.948268377	0.437488491	1.20011899	1.18E-08	1.78E-08
gLV-RA (ridge)	gLV-RA (elastic net	(-1.5, -0.2]	2.818718931	2.766408636	0	0.545899358	0.608971395
gLV-RA (ridge)	LRA	(-1.5, -0.2]	2.818718931	0.437488491	2.263061117	5.88E-12	9.58E-12
gLV-RA (elastic net	LRA	(-1.5, -0.2]	2.766408636	0.437488491	2.211341115	8.61E-12	1.38E-11

Supplemental Table 7: *p*-values associated with Figure 4F.

Module	ASV	Taxonomy
1	174	** Lachnospiraceae
1	85	* Christensenella
1	89	Bacteroides salyersiae
1	102	Ruminococcus bicirculans
1	23	Parabacteroides merdae
2	120	Holdemania massiliensis
2	39	Parabacteroides distasonis
2	141	* Mediterraneibacter
2	206	* Parabacteroides
3	121	Phocea massiliensis
3	74	* Fournierella
3	180	* Pseudoflavonifractor
3	197	* Intestinimonas
3	133	* Flavonifractor
3	49	Coprobacillus cateniformis
3	96	* Flintibacter
3	148	Blautia hydrogenotrophica
3	172	* Flintibacter
4	187	** Desulfovibrionaceae
4	128	* Butyricimonas
4	276	* Enterocloster
4	86	Coprobacter fastidiosus
4	103	** Ruminococcaceae
4	154	Alistipes indistinctus
4	165	* Lachnospira
4	169	* Dysosmobacter
4	170	* Ihubacter
5	173	** Ruminococcaceae
5	35	* Ruminococcus2 (torques group)
5	76	* Dysosmobacter
5	78	Flavonifractor plautii
5	79	* Lawsonibacter
5	192	* Anaerofilum
5	207	Intestinimonas massiliensis
5	84	** Christensenellaceae
5	227	Christensenella massiliensis
5	46	* Ruthenibacterium
5	53	* Negativibacillus
5	20	Murimonas intestini
5	151	* Lactococcus
5	156	* Flintibacter
5	109	** Lachnospiraceae
5	133	* Holdemanella
5	64	** Lachnospiraceae
5	69	Clostridium scindens
6	9	* Escherichia/Shigella
6	30	** Erysipelotrichaceae
7	153	Gemmiger formicilis
7	80	* Ruminococcus
8	261	* Ruthenibacterium
8	137	Alistipes finegoldii/onderdonkii
9	43	Parabacteroides distasonis
9	45	Parabacteroides distasonis
9	32	Parabacteroides distasonis
10	1	Bacteroides timonensis/cellulosilyticus
10	3	Bacteroides ovatus
10	4	Bacteroides fragilis/ovatus
10	7	Parasutterella excrementihominis
10	8	Phocaeicola vulgatus
10	42	* Coprobacter
10	81	** Lachnospiraceae
10	11	Bacteroides ovatus
10	132	* Frisingicoccus
10	14	Parabacteroides distasonis
10	17	* Paraprevotella
10	18	* Paraprevotella
10	145	Oxalobacter formigenes
10	160	** Lachnospiraceae
10	28	Bacteroides finegoldii

Module	ASV	Taxonomy
11	77	* Faecalibacterium
11	220	* Gemmiger
11	44	* Clostridium_IV
11	47	Blautia faecis
11	147	* Hungatella
11	152	* Acetitomaculum
12	95	Ruminococcus bromii
12	62	Bacteroides fragilis/dorei
12	124	Butyricicoccus faecihominis
12	115	* Butyricicoccus
13	2	Akkermansia muciniphila
13	177	* Caproiciproducens
13	5	* Bacteroides
13	6	Bacteroides uniformis
13	10	Parabacteroides goldsteinii
13	12	Phascolarctobacterium faecium
13	13	Bacteroides caccae
13	15	Bacteroides nordii
13	16	Bilophila wadsworthia
13	19	Phocaeicola massiliensis
13	21	* Alistipes
13	22	* Enterocloster
13	25	Lachnoclostridium urinimassiliense
13	31	Akkermansia muciniphila
13	34	* Barnesiella
13	36	Holdemania filiformis
13	37	*** Gastranaerophilales
13	38	* Phocaeicola
13	40	** Lachnospiraceae
13	41	Bacteroides uniformis
13	50	Blautia massiliensis
13	59	Butyricimonas paravirosa
13	61	Bacteroides intestinalis
13	63	Erysipelatoclostridium ramosum
13	66	Bacteroides intestinalis
13	67	Butyricimonas faecihominis
13	75	* Faecalicatena
13	88	* Butyricimonas
13	123	Odoribacter splanchnicus
13	129	* Monoglobus
13	130	* Blautia
13	136	Alistipes finegoldii
13	144	* Blautia
13	157	** Lachnospiraceae
13	166	* Lawsonibacter
14	24	Blautia wexlerae/obeum
14	26	Hungatella hathewayi/effluvii
14	29	** Lachnospiraceae
14	33	Eisenbergiella massiliensis
14	48	** Lachnospiraceae
14	52	Blautia caecimuris
14	54	** Lachnospiraceae
14	57	** Lachnospiraceae
14	58	* Hungatella
14	60	Sellimonas intestinalis
14	83	** Lachnospiraceae
14	94	* Blautia
14	101	** Lachnospiraceae
14	105	* Enterocloster
14	107	* Enterocloster
14	108	*** Clostridiales
14	111	Anaerostipes caccae
14	117	*** Clostridiales
14	122	Monoglobus pectinilyticus
14	134	* Sellimonas
14	140	* Blautia
14	171	** Lachnospiraceae
15	65	* Holdemania
16	144	* Blautia
16	127	* Paraprevotella
16	118	* Paraprevotella
17	70	** Lachnospiraceae

Supplemental Table 8: Module members. Taxa comprising each module along with ASV number.