BRIGHAM AND WOMEN'S HOSPITAL

Robust and Scalable Models of Microbiome Dynamics

Travis E. Gibson and Georg K. Gerber

Massachusetts Host Microbiome Center, Brigham and Women's Hospital, Harvard Medical School

tgibson@mit.edu

@travisegibson

HARVARD MEDICAL SCHOOL

Funding: DARPA BRICS award HR0011-15-C-0094, NIH T32

Introduction

Microbes are everywhere including in and on our bodies, and have been shown to play key roles in a variety of prevalent human diseases. Consequently, there has been intense interest in the design of bacteriotherapies or "bugs as drugs", which are communities of bacteria administered to patients for specific therapeutic applications. Central to the design of such cocktails is the knowledge (or inference) of a causal microbial interaction network and prediction of the population dynamics of the organisms. In this work we present a Bayesian nonparametric model and associated efficient inference algorithm that addresses the key conceptual and practical challenges of learning microbial dynamics from time series microbe abundance data.

Interpretability and the Microbiome

Challenges associated with inference in the microbiome:

- High-dimensional (300+ strains of bacteria in the gut, **potentially 100,000 microbe-microbe interactions**).
- Temporally sparse and non-uniformly sampled data.
- High measurement noise
- Nonlinear and physically non-negative dynamics.

Many potential interactions: how do we decide what interactions are "real" and how do we simplify the interaction landscape so as to be interpretable?

Contribution:

1. Introduction of a temporally varying **auxiliary variable** technique to enable efficient inference by relaxing hard non-negativity constraint
2. Clustering of microbes into redundant **interaction modules** and **structural learning** of a compact interaction network among modules.

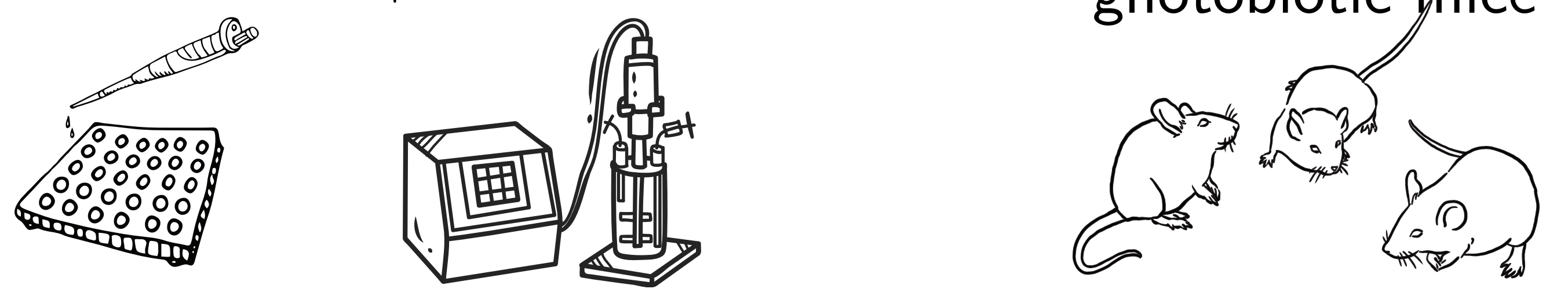
Experimental Design for Inference

In vitro

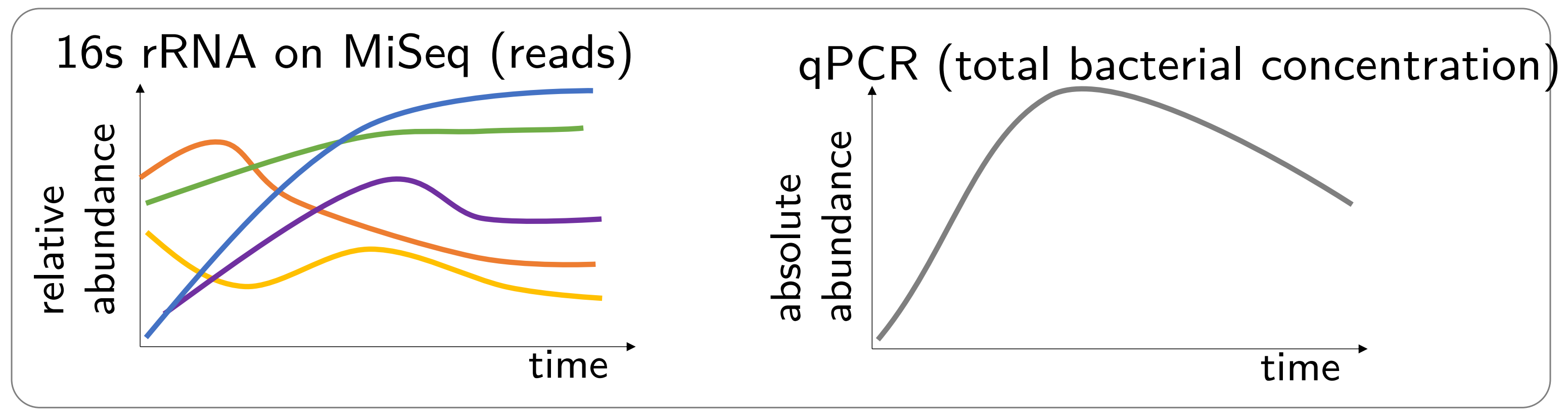
batch + chemostat

In vivo

gnotobiotic mice

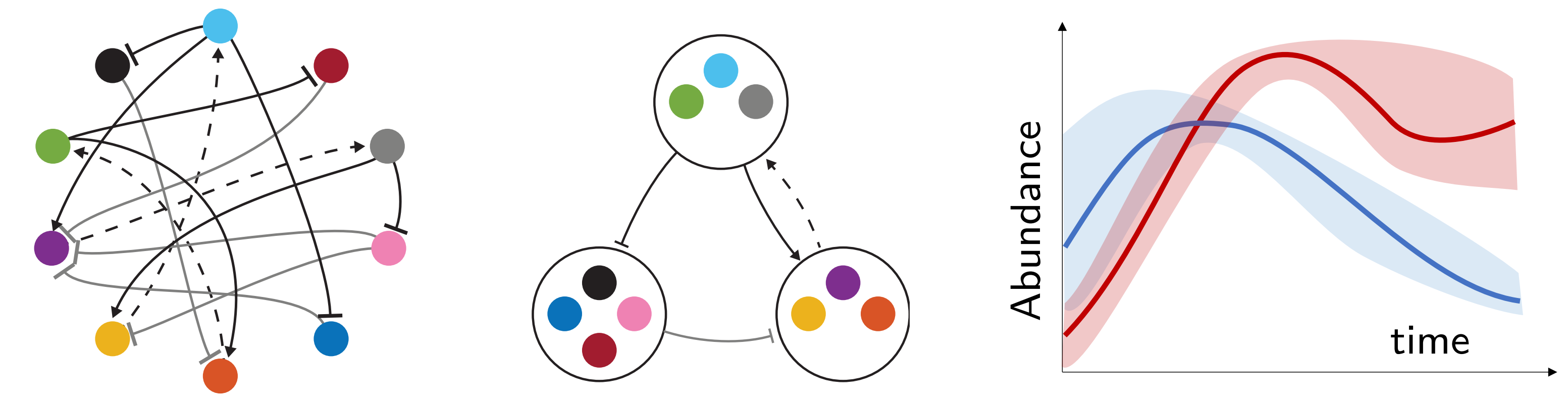


Perturbations / Biological Replicates



+

Bayesian Inference



Inferred Network

Interaction Modules

Forecast Abundances

Dynamical Model

Lotka-Volterra Dynamics Abundance of Microbe i at time t : $\mathbf{x}_{t,i}$

$$\frac{d\mathbf{x}_{t,i}}{dt} = \underset{\text{growth parameter}}{\alpha_i}\mathbf{x}_{t,i} + \underset{\text{self limiting}}{\beta_{ii}}\mathbf{x}_{t,i}^2 + \sum_{j\neq i}\underset{\text{interaction coefficient}}{\beta_{ij}}\mathbf{x}_{t,i}\mathbf{x}_{t,j} + \underset{\text{disturbance term}}{\frac{d\mathbf{w}_{t,i}}{dt}}$$

Discrete approximation to the Lotka-Volterra dynamics

$$\mathbf{x}_{k+1,i} = \mathbf{x}_{k,i} + \left(\alpha_i\mathbf{x}_{k,i} + \beta_{ii}\mathbf{x}_{k,i}^2 + \sum_{j\neq i}\beta_{ij}\mathbf{x}_{k,i}\mathbf{x}_{k,j}\right)\Delta_k + (\mathbf{w}_{k+1,i} - \mathbf{w}_{k,i})\sqrt{\Delta_k}$$

k : discrete time index

discrete time step size

- Cluster the microbes into interaction modules (**Dirichlet Process** prior)
- No interactions **within** cluster, only **between** clusters.
- **Edge Selection**: add indicator variables for cluster-cluster interactions (Bayesian variable selection, structure learning for graphical models)
- Insert auxiliary variable $\mathbf{q}$ between the measurements $\mathbf{y}$, $\mathbf{Q}$ and the state $\mathbf{x}$

Comment: not allowing within cluster interactions dramatically reduces the number of inferred interactions and is consistent with the biologically important scenario of redundant functionality among sets of microbes

Additional Model Components

Dirichlet Process (clustering)

cluster assignment

$$\boldsymbol{\pi}_\mathbf{c} \mid \alpha \sim \texttt{Stick}(\alpha)$$

$$\mathbf{c}_i \mid \boldsymbol{\pi}_\mathbf{c} \sim \texttt{Multinomial}(\boldsymbol{\pi}_\mathbf{c})$$

$$\mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j} \mid \boldsymbol{\sigma}_\mathbf{b} \sim \texttt{Normal}(0, \boldsymbol{\sigma}_\mathbf{b}^2)$$

cluster interactions coefficient

Edge Selection

$$\mathbf{z}_{\mathbf{c}_i,\mathbf{c}_j} \mid \boldsymbol{\pi}_\mathbf{z} \sim \texttt{Bernouli}(\boldsymbol{\pi}_\mathbf{z})$$

indicator variable for cluster interaction

Self Interaction

$$\mathbf{a}_{i,1}, \mathbf{a}_{i,2} \mid \boldsymbol{\sigma}_\mathbf{a} \sim \texttt{Normal}(0, \boldsymbol{\sigma}_\mathbf{a}^2)$$

Dynamics

$$\mathbf{x}_{k+1,i} \mid \mathbf{x}_k, \mathbf{a}_i, \mathbf{b}, \mathbf{c}, \mathbf{z}, \boldsymbol{\sigma}_\mathbf{w} \sim \texttt{Normal}\left(\mathbf{x}_{k,i} + \mathbf{x}_{k,i}\left(\mathbf{a}_{i,1} + \mathbf{a}_{i,2}\mathbf{x}_{k,i} + \sum_{\mathbf{c}_j\neq\mathbf{c}_i}\mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j}\mathbf{z}_{\mathbf{c}_i,\mathbf{c}_j}\mathbf{x}_{k,j}\right), \Delta_k\boldsymbol{\sigma}_\mathbf{w}^2\right)$$

Measurement Model

auxiliary variable $\quad \mathbf{q}_{k,i} \mid \mathbf{x}_{k,i} \sim \texttt{Normal}(\mathbf{x}_{k,i}, \boldsymbol{\sigma}_\mathbf{q}^2) \qquad \mathbf{q}_{k,i} \sim \texttt{Uniform}[0, \texttt{L})$

reads $\quad \mathbf{y}_{k,i} \mid \mathbf{q}_{k,i} \sim \texttt{NegBin}(\phi(\mathbf{q}_k), \epsilon(\mathbf{q}_k))$

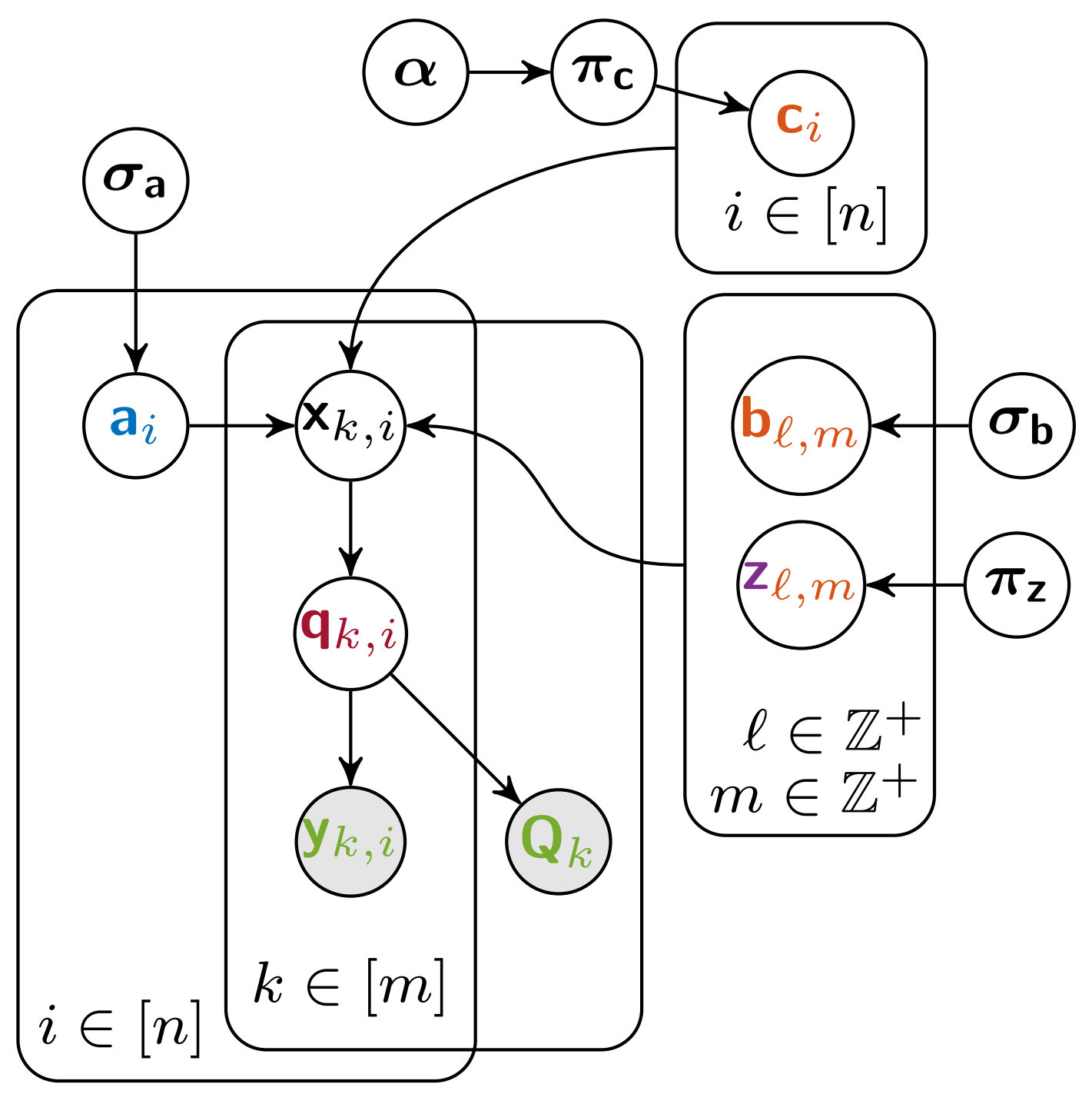
qPCR $\quad \mathbf{Q}_k \mid \mathbf{q}_{k,i} \sim \texttt{Normal}\left(\sum_i \mathbf{q}_{k,i}, \sigma_{\mathbf{Q}_k}^2\right)$

Introduction of $\mathbf{q}$ allows for

- Efficient Gibbs/collapsed Gibbs sampling
- Posterior distributions for coefficients $\mathbf{a}$, $\mathbf{b}$ are Gaussian, (direct sampling from posterior)
- Can Marginalize out in closed form the interaction coefficients $\mathbf{b}$ for both module learning and structure learning

Without $\mathbf{q}$, dynamics would have been modeled with truncated distribution, resulting in the posteriors of $\mathbf{a}$, $\mathbf{b}$ being truncated as well and not allowing for marginalization elsewhere in the model.

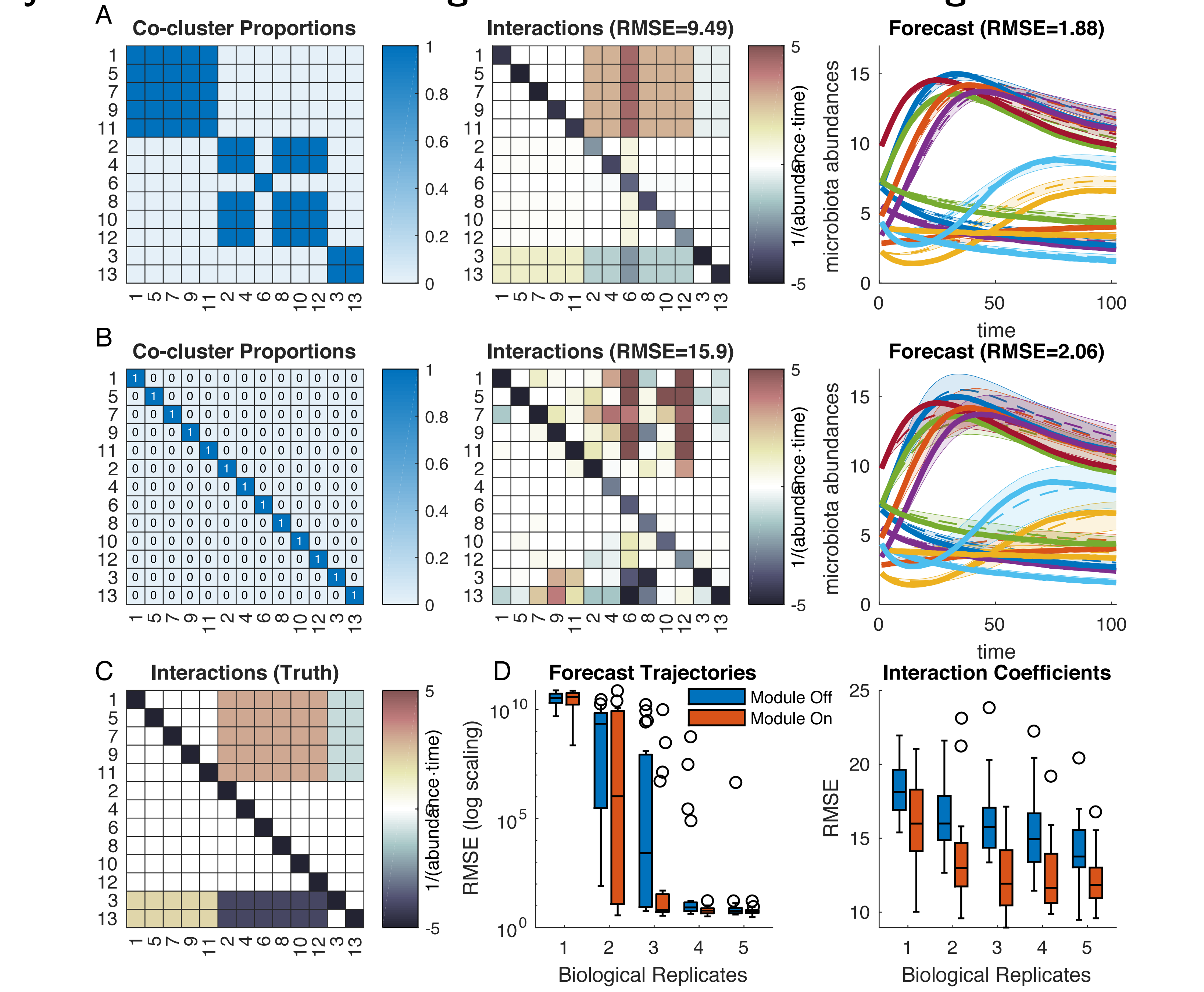
Graphical Model



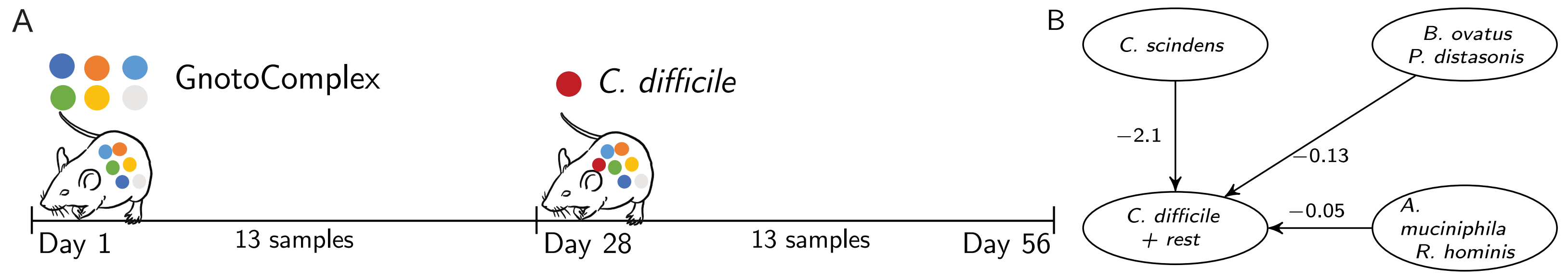
[1] Bucci, Vanni, et al. "MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses." *Genome Biology* (2016)

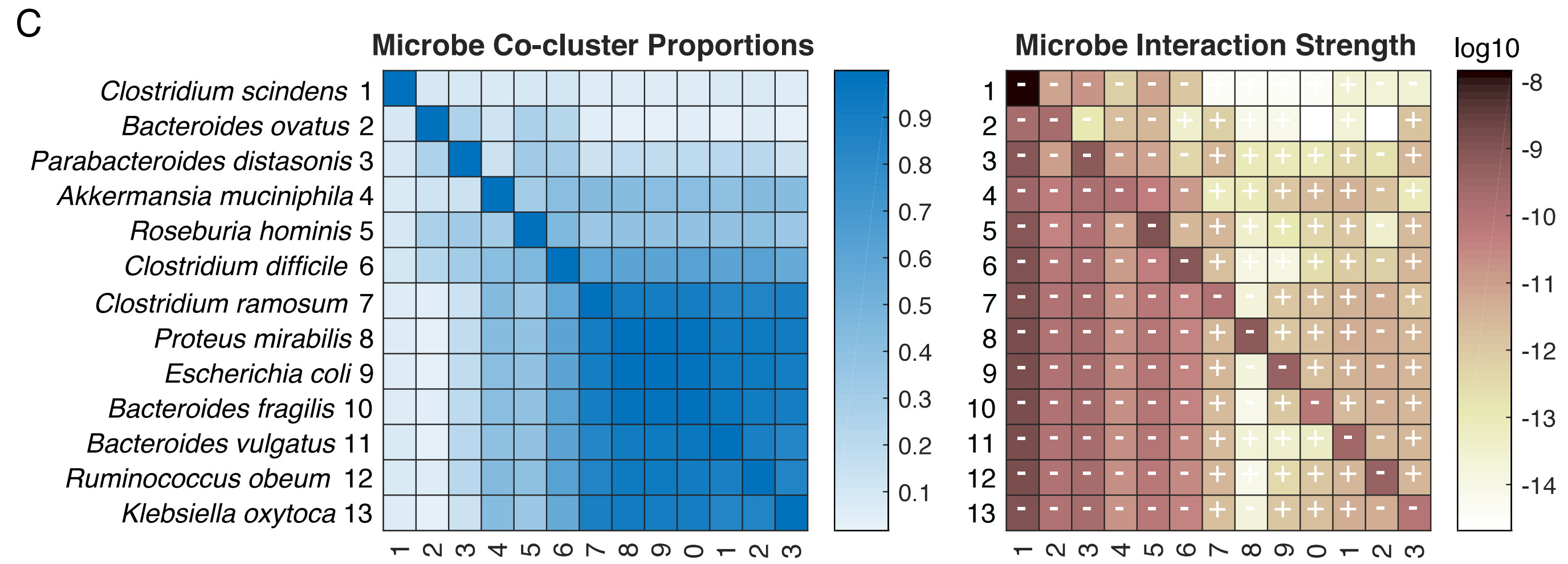
Inference on Synthetic and In Vivo Time Series

Synthetic data illustrating the use of module learning



Learning interaction modules from *in vivo* time series [1]

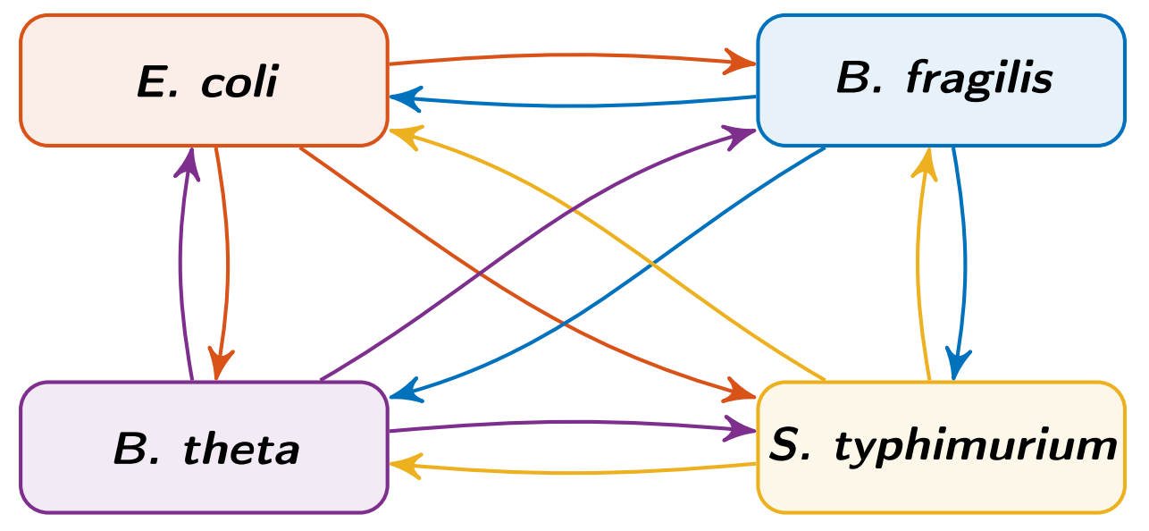




Ongoing Projects

Microbial Interaction Engineering

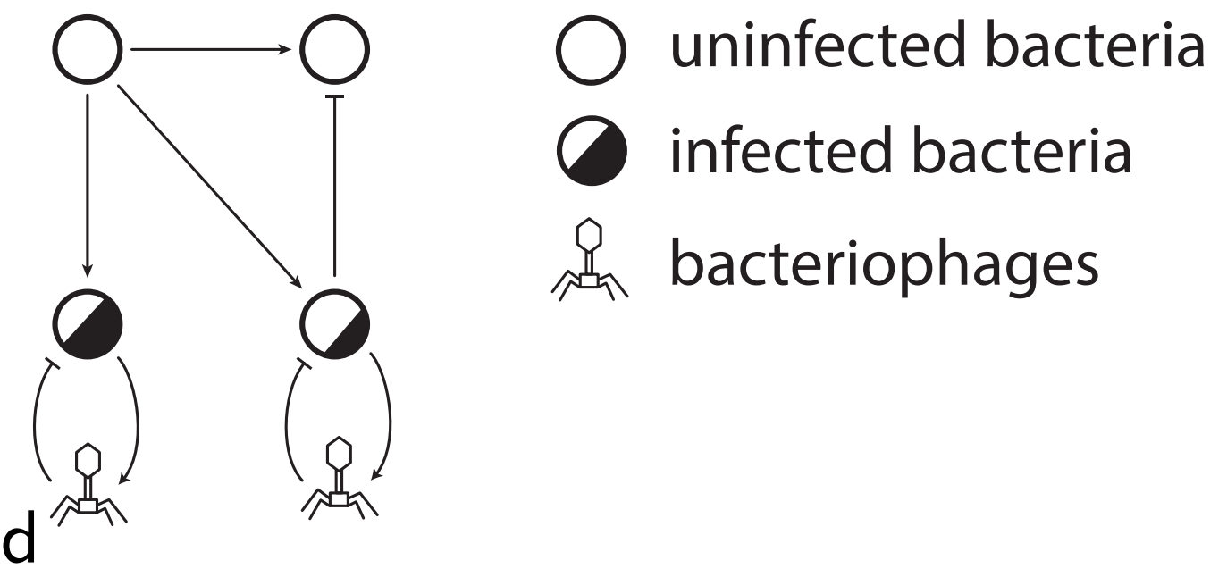
Marika Ziesack
Silver Lab, Harvard



- Engineered to overproduce 1 amino acid
- Engineered to need 3 amino acids

Bacteriophage Therapy

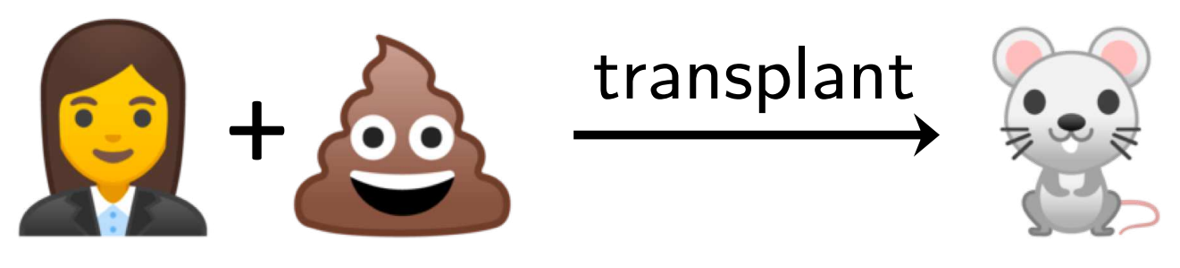
Bryan Hsu
Silver Lab, Harvard



- Bacteriophages are "bacteria viruses"
- Phages as control knobs to modulate microbiome

Complex Microbiota

Massachusetts
Host Microbiome
Center



- Colonize mice with human fecal sample containing 300+ bacteria