Submitted ACC 2016

The Human Microbiome: Opportunities for Dynamics, Systems, and Control (based on the IFAC blog post *Translate or Die*)

Travis E. Gibson

***Abstract*—Yu-Chi "Larry" Ho famously titled a 2010 blog post "Control Is Dead?",**
http://blog.sciencenet.cn/blog-1565-344686.html
In follow up posts Larry's advice has been simple: control has to be constantly reinvented with the times. It's success is its own worst enemy. Therefore, control engineers and theorists must continue to push the boundaries of possible applications and take advantage of the adage "necessity is the mother of invention." This paper gives a brief outline of a new research area in biology and medicine where problems for systems, dynamics, and control are sitting and waiting for solutions and insights. How much do you know about microbes?

I. Introduction

They are everywhere. Some 100 trillion inhabit the earth, comprising half of the animal mass on it. Have you guessed what I am talking about yet? See the following articles in the New York Times [25], NY Times Magazine [29], Scientific American [20], Nature [1], Science [28], or this TED Talk [22] with the accompanying book [21] to refresh your memory. Now the human microbiome has been associated with almost every disease possible, microbes in the gut have even been associated with brain diseases [27]. The study of these little things is kind of a big deal.

What is a normal human microbiome?

The most important developments in the human microbiome have come via the analysis of large cohorts across body sites (gut, mouth, vagina, skin, etc) [34] and longitudinal studies where fecal samples have been collected on a daily scale [12, 14]. What we know from these studies is that the abundance and kinds of microbes are body site specific. Figure 1 illustrates this point.

In Figure 1 the relative abundance of microbes for 4,788 specimens from 242 adults are projected onto the first two principle coordinates, see Appendix A for a discussion about principle coordinates.[1] The different body sites are color coded, and it is clear that the specimens cluster according to body site and not by subject. We have also learned that microbial abundances are fairly stable for each site and for each subject (I will discuss this in more detail shortly). Before getting to the dynamics and estimation part we need a story so as to understand the translational implications of a better understanding of the human microbiome.

Fecal Microbial Transplantation

This story begins with Jane coming to the hospital because of an infection in her leg. To kill the infection she is given broad spectrum antibiotics. After a few days the infection is gone, but Jane now has severe diarrhea. The antibiotics have killed some of the healthy bacteria in her gut and now Jane has an over abundance of *Clostridium difficile*, i.e. she has *Clostridium Difficile Infection* (CDI). Ironically the most often prescribed treatment for CDI is another antibiotic. This targeted antibiotic always works in temporarily reducing the abundance *C. difficile*, but the CDI is recurrent. So with no other options Jane asks her brother John for a fecal sample. This fecal sample is prepared and transplanted into Jane (*Fecal Microbial Transplantation* (FMT)). As if a miracle has occurred Jane is healthy again. This kind of story is becoming common place in hospitals around the country now [31].

What happens in terms of the abundance of the microbes post-FMT is quite amazing. Figure 2 shows the trajectory gut microbes take before and after an FMT. Several subjects stool samples pre-FMT are circled in red and the trajectories (post-FMT) are seen to rapidly converge to the green circle (which also contains the host sample), overlaid on top of samples from the 242 healthy adults from Figure 1. A movie of these trajectories can be downloaded here [2]. While the post-FMT samples do deviate slightly from the host sample in terms of relative abundance, the samples remain within the range

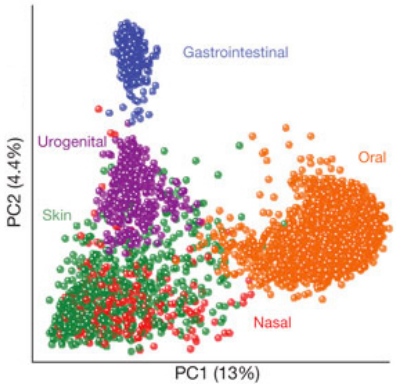

Fig. 1. Principle coordinates of samples illustrating variation of the microbes between body sites. Adapted from [34, Figure 1c]. Permission from Nature Publishing Group, license number 3700220673812 Copyright Clearance Center.

T. E. Gibson is with Harvard Medical School and the Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston MA, 02115, email: (tgibson@mit.edu)

[1] In the original article it is not made clear as to what phylogenetic depth the microbes are organized for this figure. It is assumed from context that the microbes are organized in terms of *Operational Taxonomic Units* (OTUs), which for simplicity one can think of as the taxonomic rank of species. More information on OTUs is given in a later section.

Fig. 2. Trajectories of fecal samples in principle coordinates for patients receiving FMT. Pre FMT circled in red. Post FMT 'steady state' circled in green. The lines are the trajectories of the daily stool samples projected into the principle coordinates. Adapted from [37, Figure 1]. Creative Commons Attribution (CC BY) license.

of what is considered to be healthy. Stated simply what we are observing is the patients gut microbiome reconstituted and remaining in an abundance profile similar to that of the donor. It is quite amazing.

II. Is the microbiome stable?

What we just saw above was that the post FMT stool samples remained similar to the host after transplantation. So then one natural question arises: How stable is the human microbiome? Biologist recognize that this is an important subject as is evidenced by Figure 3 which appeared in a recent review article in Science [13].

We should be delighted to see that the notion of stability has been recognized as an important issue in the human microbiome. There appears to be a misunderstanding of what the word stable means however. This is simply an ignorance issue and as control engineers/theorists we should just simply educate those in this field. Consider Figure 4 that shows 15 days of samples (shown in yellow) taken from the one year gut microbiome study in [12], and projected onto the principle coordinates from a previous study [3,4]. Ignore the red, green, and blue dots and focus on the trajectories of the yellow dots with gray lines following the day to day changes in the stool samples. The authors of [23] wanted to highlight the fact that the samples can deviate from steady state in almost all directions. The authors unfortunately draw the conclusion that this is a visualization of instability in the gut microbiome. The original figure from the study in [23] is shown on the left and the annotated figure is on the right. Note that the two trajectories after deviation return to the "steady region". This is not instability, but the very definition of stability. One could even argue that we are observing asymptotic like stability, i.e. in the absence of disturbances all trajectories converge to a single fixed point. Unfortunately biological systems are noisy at the input, output, and in terms of model parameters. Could this line of reasoning help to explain the success of FMT? I think you can begin to see

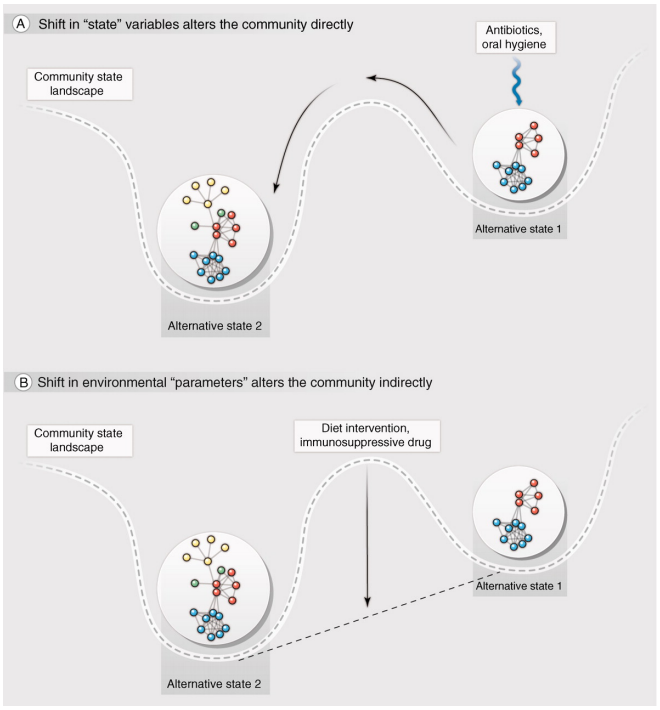

Fig. 3. An illustration of the stability landscape for a microbial ecosystem. The top figure illustrates that through the introduction of 'state' disturbances a population can be pushed into another region of local stability. The bottom figure illustrates that the 'state' of an ecosystem can also be alterred indirectly by changing environmental parameters (an example of this in the gut microbiome would be an extreme diet change). Adapted from [13, Figure 1]. Permission from the American Association for the Advancement of Science, license number 3700221448858 Copyright Clearance Center.

where those working in the area of dynamics and control might be needed in this emerging field.

III. How do we model the microbiome?

The most common way that microbes interact is through the consumption of nutrients and the synthesis of products (not necessarily through the direct consumption of each other) [26]. Therefore, a detailed model would contain states for both the abundance of microbes and the abundance of the metabolites they consume and synthesize. At the finest level of modeling all host and microbe metabolic pathways would need to be mapped. We currently do not poses the technology or sufficiently rich data to perform this rigorously. At this point in our understanding of microbial dynamics it is more common to think of a reduced order model that only accounts for the abundances of the microbes.

The two most popular (reduced order) models are *Generalized Lotka-Volterra* (GLV) dynamics over a network and Bayesian networks [16]. The first is deterministic and is the most common one studied in the literature, while the second is probabilistic. I will focus on the first one here, but a similar discussion could follow with a probabilistic mind set as well.

Let x_i be the abundance of microbe i for subject 1 at a specific location on/in the body. Let's assume for now we are concerned only with the gut. Then the GLV model for n microbes interacting in the gut of subject 1 is described by

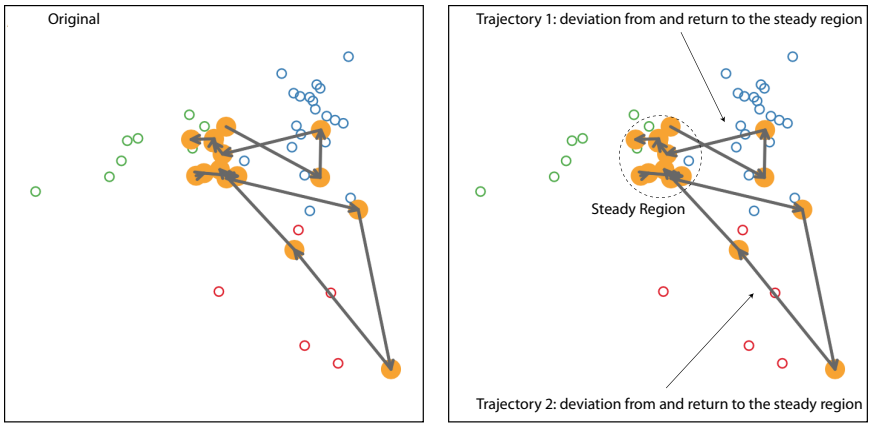

Fig. 4. Two trajectories of consecutive daily samples from the gut of the male subject in [12]. Adapted from [23, Figure 3A]. Fair use, Copyright Act of 1976, 17 U.S.C. §107.

the following differential equation

$$\dot{x}_i = r_i x_i + \sum_{j=1}^{n} a_{ij} x_i x_j$$

where $i = 1, 2, ..., n$. Collecting the abundances of the microbes into a column vector $x = [x_1,\ x_2,\ \ldots,\ x_n]^{\mathsf{T}}$ the dynamics can be compactly written as

$$\dot{x} = \mathsf{diag}(r)x + \mathsf{diag}(x)Ax, \tag{1}$$

where diag takes a column vector and returns a matrix with the column vector along the diagonal, r is a column vector of the r_i and $[A]_{ij} \triangleq a_{ij}$. In this modeling paradigm r captures linear growth or death and the matrix A captures causal interactions amongst species, we will refer to A as the microbial interaction matrix or simply interaction network. Thus, a_{ij} represents the average affect that species j has on species i by determining what species j generates as products and what both species i and j consume as nutrients. For instance, if species j produces products that species i consumes as nutrients and they do not compete for any other nutrients then a_{ij} would be positive. For a short discussion on a sufficient condition for stability of the above dynamics see Appendix B.

As previously mentioned we dot not fully understand the microbial-metabolic interactions well enough to have a global bottom up model. Do we have sufficient data to learn the interactions in the simplified GLV model? We will discuss this in more detail shortly.

Now consider the gut of a different individual, subject 2, and assume that the dynamics are as follows

$$\dot{y} = \mathsf{diag}(\bar{r})y + \mathsf{diag}(y)\bar{A}y.$$

Notice that I have written the dynamics for both subjects with different variables. The couple (r, A) represents the growth rates and interaction matrix for subject 1 and the couple $(\bar{r}, \bar{A})$ for subject 2. Is it possible that for two otherwise healthy individuals with similar diet $A = \bar{A}$ and $r = \bar{r}$. Recent attempts to infer the interaction matrices for two individuals illustrates some short comings in the literature and another opportunity for those working in system identification and machine learning to have an immediate impact in this field.

A. Some Comments on Microbiome Network Reconstruction

As eluded to earlier we are interested in reconstructing the microbial interaction network for an individual (or all individuals) and most importantly we would like to know if two individuals will have the same network. In this section we will first transform the dynamics in (1) so that linear regression can be performed [6, 7, 32]. Beginning with the dynamics in (1) and dividing by x_i we have that

$$\dot{x}_i/x_i = r_i + \sum_{j=1}^{n} a_{ij} x_j, \quad x_i \neq 0.$$

Integrating from t_k to t_{k+1} and using the approximation that $x(t)$ is constant over $t \in [t_k, t_{k+1})$ we have that

$$\log(x_i(t_{k+1})) - \log(x_i(t_k)) = \Delta_k \left(r_i + \sum_{j=1}^{n} a_{ij} x_j(t_k) \right) + e_i(t_k). \tag{2}$$

The error term e_i arrises from the fact that this is an approximation.[2]

Equation (2) can be rewritten in terms of a regressor vector

$$\phi(k) = [1, x_1(t_k),\ x_2(t_k),\ \ldots,\ x_n(t_k)]^{\mathsf{T}},$$

the parameter vector $\theta_i = [r_i,\ a_{i1},\ a_{i2},\ \ldots,\ a_{in}]$, and the log difference $y_i(k) = \log\left(x_i(t_{k+1})\right) - \log\left(x_i(t_k)\right)$ as

$$e_i(k) + y_i(k) = \theta_i \phi(k).$$

[2]We note that with little effort the integration of $\sum_{j=1}^{n} a_{ij} x_j$ could be improved given the information available. The approximation in (2) is used only because it is the one most frequently used in the literature. That being said, an approximation of the term $\int_{t_k}^{t_{k+1}} x_j(\tau)\, \mathrm{d}\tau$, using information about x_j at both endpoints, would result in $\frac{\Delta_k}{2}(x_j(t_k) + x_j(t_{k+1}))$, and using this information (2) could be replaced by $\log(x_i(t_{k+1})) - \log(x_i(t_k)) = \Delta_k(r_i + \frac{1}{2}\sum_{j=1}^{n} a_{ij}(x_j(t_k) + x_j(t_{k+1})) + \tilde{e}_i(t_k)$ with a new error term $\tilde{e}_i$.

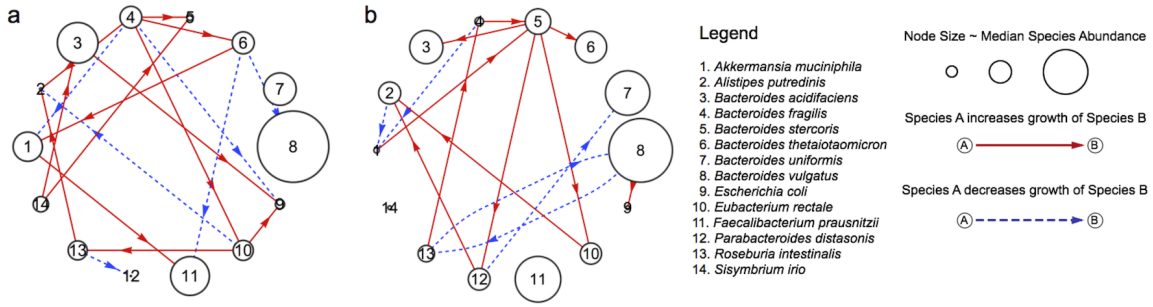

Fig. 5. Inferred interaction subnetworks of the gut microbiome for the 14 most abundant species from the two subjects in the longitudinal data presented in [12]. Adapted from [17, Figure 6]. Creative Commons Attribution (CC BY) license.

The identification problem can then be defined as finding the parameter matrix estimate $\hat{\Theta} = [\hat{\theta}_1^\mathsf{T}, \hat{\theta}_2^\mathsf{T}, \cdots, \hat{\theta}_n^\mathsf{T}]^\mathsf{T}$ of the true parameter matrix $\Theta = \left[\theta_1^\mathsf{T}, \theta_2^\mathsf{T}, \cdots, \theta_n^\mathsf{T}\right]^\mathsf{T}$. Letting

$$y(k) = [y_1(k),\ y_2(k),\ \ldots, y_n(k)]^\mathsf{T}$$

be the log difference vector for all species and $Y = [y(1),\ y(2),\ \ldots,\ y(N-1)]$ be the log difference matrix, the system identification problem can be compactly presented as

$$\min_{\hat{\Theta}} \|Y - \hat{\Theta}\Phi\|_\mathsf{F}^2$$

where $\Phi = [\phi(1),\ \phi(2),\ \ldots,\ \phi(N-1)]$ is the regressor matrix and $\|\cdot\|_\mathsf{F}$ denotes the Frobenius norm. If $\Phi\Phi^\mathsf{T}$ is full rank then the solution to the above minimization problem is $Y\Phi^\mathsf{T}(\Phi\Phi^\mathsf{T})^{-1}$. This however may not be the case and so one often introduces regularization into the problem so as to overcome the rank deficiency of the regressor product $\Phi\Phi^\mathsf{T}$. The L2 regularized optimization problem is defined as

$$\min_{\hat{\Theta}} \left(\|Y - \hat{\Theta}\Phi\|_\mathsf{F}^2 + \lambda \|\hat{\Theta}\|_\mathsf{F}^2 \right)$$

where $\lambda > 0$ is the Tikhonov regularization term [36].[3] The minimal solution to the above problem can be given directly as

$$\arg\min_{\hat{\Theta}} \left(\|Y - \hat{\Theta}\Phi\|_\mathsf{F}^2 + \lambda \|\hat{\Theta}\|_\mathsf{F}^2 \right) = Y\Phi^\mathsf{T}(\Phi\Phi^\mathsf{T} + \lambda I)^{-1}$$

where I is the identity matrix.[4] Note that $\Phi\Phi^\mathsf{T} + \lambda I$ is full rank for any $\lambda > 0$.

The linear regression technique just described was applied to the data collected from [12] (daily stool samples from two individuals over one year) with the identified subnetwork for the most abundant species shown in Figure 5 [17]. Just by inspection one can see that the two networks are different. Thus the conclusion is that the same microbes interact differently in their hosts. This result actually goes against our intuition. Given that the metabolic pathways for otherwise healthy individuals are the same, microbes should on average have host independent dynamics, otherwise treatments like FMT would not work and furthermore post-FMT patients would not reconstitute into the abundance profile of the donor. There are two issues with the analysis in [17].

First, only *relative species abundance* is available from any given sample. Let us briefly discuss how microbe abundance profiles are generated from a sample [19]. Once a sample is prepared from a subject it is either sequenced with a marker gene or shotgun sequencing is performed. After sequencing and quality measures are taken, each gene of interest in a sample has a read count. The most common way to discuss microbe abundance is to then group gene reads based upon *Operational Taxonomic Units* (OTUs), sometimes called phylotypes [19, page 255]. Given that there are variations in sample size, OTU abundances between samples are always compared in terms of relative abundance and not absolute reads.

This is an issue for system identification, as will be explored now. Let the relative abundance of microbe i for a given sample be defined as

$$\tilde{x}_i = \frac{x_i}{x_\Sigma}, \quad x_\Sigma = \sum_{i=1}^n x_i.$$

Then, rewriting the dynamics in (2) using relative abundance we have

$$\log\left(\frac{x_\Sigma(t_{k+1})}{x_\Sigma(t_k)}\frac{\tilde{x}_i(t_{k+1})}{\tilde{x}_i(t_k)}\right) = \Delta_k\left(r_i + x_\Sigma(t_k)\sum_{j=1}^n a_{ij}\tilde{x}_j(t_k)\right) + e_i(t_k).$$

So as to overcome the fact that $x_\Sigma(t_k)$ is unknown at any time point t_k, in [17] it is assumed that $x_\Sigma(t_k)$ is a constant for all t_k. This assumption most likely introduced significant error in the network reconstruction. Inferring dynamics from relative abundance measurements alone is not feasible.

Second, given that the data is not sufficiently rich the authors of [17] decided to only use the union of the top 10 most abundant species from each subject (resulting in the 14 species shown in Figure 5) in the linear regression analysis, instead of using regularization techniques. At the

[3]The Frobenius norm is used here and not the induced 2-norm as the original problem definition for linear reression is given in terms of column vectors a, b and a matrix A such that the $\|Ab - a\|_2^2 + \lambda\|b\|_2^2$ is minimized. The Frobenius norm is a more natural extension to the multilinear optimization problem as it is computationally more straight forward than the induced 2-norm.

[4]Note that there has been significant interest in L1 "lasso" regularization [35] do to results by Tao, Candès, and Donaho [8–11, 15].

species taxonomic rank there are on the order of 200 species in a human stool sample. Thus in the generation of Figure 5 more than 90% of the data was not used. It is because of these two major issues that the conclusions drawn from the analysis in [17] are most likely not correct. Unfortunately others have unknowingly built upon this flawed analysis [33].

A similar work looked to infer a subnetwork of microbial interactions [7], but here the authors focused on those species that interact most strongly with *C. difficile*. This work overcomes the issue of relative abundance by having recorded the total microbe DNA per gram for each sample. In addition L2 regularization is used with the Tikhonov regularization parameter chosen using cross-validation [24]. The authors however do make the mistake of only using the most abundant species in the linear regression analysis. Needless to say the results from this work are much more reliable and the authors are able to confidently illustrate clinically supported findings from the inferred subnetwork, see Figure 6.

System identification in biological networks (sometimes referred to as network reconstruction) needs the influence of control engineers [5]. In addition to some of the issues just raised lots of open question remain:

- Are some body sites more stable than others?
- How do we rigorously demonstrate this stability?
- Are the networks of two healthy individuals similar?
- How do different diseases affect that network?
- Why do FMTs work?
- Are there other modeling approaches that can be used to understand microbial dynamics?
- What are the fundamental limitations for network reconstruction?
- Finally, how do we control the microbiome?

IV. Conclusions

Aircraft control has been one of the cornerstone applications for control for more than 50 years. It is time however to find new areas for research. I hope this has inspired you to consider the human microbiome as a possible research area for the application of everything you have learned in dynamics, control, and system identification.

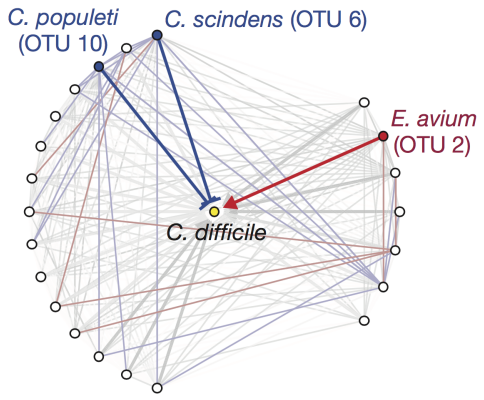

Fig. 6. Inferred subnetwork of microbes common to both human and mouse that most strongly interact with *C. difficile*. The authors then went on to show that indeed *C. scindens* could in fact ameliorate *C. difficile* infection, see [7, Figure 3]. Adapted from [7, Figure 2f] Permission from Nature Publishing Group, license number 3718211295678 Copyright Clearance Center.

V. Acknowledgements

I would like to acknowledge my collaborators Yang-Yu Liu and Amir Bashan, as well as conversations I have had with Aimee Milliken, Eric Alm, Curtis Huttenhower, and Rob Knight.

References

[1] available at http://www.nature.com/nature/focus/humanmicrobiota/.
[2] available at http://www.microbiomejournal.com/content/download/supplementary/s40168-015-0070-0-s1.mp4.
[3] Manimozhiyan Arumugam et al., *Enterotypes of the human gut microbiome*, Nature **473** (2011), no. 7346, 174–180.
[4] _____, *Addendum: Enterotypes of the human gut microbiome*, Nature **506** (2014), no. 7489, 516–516.
[5] Philippe Bastiaens, Marc R Birtwistle, Nils Bluthgen, Frank J Bruggeman, Kwang-Hyun Cho, Carlo Cosentino, Alberto de la Fuente, Jan B Hoek, Anatoly Kiyatkin, Steffen Klamt, Walter Kolch, Stefan Legewie, Pedro Mendes, Takashi Naka, Tapesh Santra, Eduardo Sontag, Hans V Westerhoff, and Boris N Kholodenko, *Silence on the relevant literature and errors in implementation*, Nature Biotechnology **33** (2015), no. 4, 336–339.
[6] Vanni Bucci and Joao B. Xavier, *Towards predictive models of the human gut microbiome*, Journal of Molecular Biology **426** (2014), no. 23, 3907–3916.
[7] Charlie G. Buffie, Vanni Bucci, Richard R. Stein, Peter T. McKenney, Lilan Ling, Asia Gobourne, Daniel No, Hui Liu, Melissa Kinnebrew, Agnes Viale, Eric Littmann, Marcel R. M. van den Brink, Robert R. Jenq, Ying Taur, Chris Sander, Justin R. Cross, Nora C. Toussaint, Joao B. Xavier, and Eric G. Pamer, *Precision microbiome reconstitution restores bile acid mediated resistance to clostridium difficile*, Nature **517** (2015), no. 7533, 205–208.
[8] Emmanuel J Candès, Justin Romberg, and Terence Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, Information Theory, IEEE Transactions on **52** (2006), no. 2, 489–509.
[9] Emmanuel J Candes, Justin K Romberg, and Terence Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on pure and applied mathematics **59** (2006), no. 8, 1207–1223.
[10] Emmanuel J Candes and Terence Tao, *Decoding by linear programming*, Information Theory, IEEE Transactions on **51** (2005), no. 12, 4203–4215.
[11] _____, *Near-optimal signal recovery from random projections: Universal encoding strategies?*, Information Theory, IEEE Transactions on **52** (2006), no. 12, 5406–5425.
[12] J Gregory Caporaso, Christian L Lauber, Elizabeth K Costello, Donna Berg-Lyons, Antonio Gonzalez, Jesse Stombaugh, Dan Knights, Pawel Gajer, Jacques Ravel, and Noah Fierer, *Moving pictures of the human microbiome*, Genome Biology (2011).
[13] Elizabeth K. Costello, Keaton Stagaman, Les Dethlefsen, Brendan J. M. Bohannan, and David A. Relman, *The application of ecological theory toward an understanding of the human microbiome*, Science **336** (2012), no. 6086, 1255–1262.
[14] Lawrence A David, Arne C Materna, Jonathan Friedman, Maria I Campos-Baptista, Matthew C Blackburn, Allison Perrotta, Susan E Erdman, and Eric J Alm, *Host lifestyle affects human microbiota on daily timescales*, Genome Biology (2014).
[15] David L Donoho, *Compressed sensing*, Information Theory, IEEE Transactions on **52** (2006), no. 4, 1289–1306.
[16] Karoline Faust and Jeroen Raes, *Microbial interactions: from networks to models*, Nature Reviews Microbiology **10** (2012), no. 8, 538–550.
[17] Charles K. Fisher and Pankaj Mehta, *Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression*, PLoS ONE **9** (2014), no. 7, e102451.
[18] B. S. Goh, *Global stability in many-species systems*, American Naturalist (1977), 135–143.
[19] Julia K. Goodrich, Sara C. Di Rienzi, Angela C. Poole, Omry Koren, William A. Walters, J. Gregory Caporaso, Rob Knight, and Ruth E. Ley, *Conducting a microbiome study*, Cell **158** (20147), no. 2, 250–262.

[20] Christine Gorman, *Explore the human microbiome*, Scientific American (May 15 2012), available at http://www.scientificamerican.com/article/microbiome-graphic-explore-human-microbiome/.
[21] Rob Knight, *Follow your gut: The enormous impact of tiny microbes*, Simon and Schuster, 2015.
[22] ———, *How our microbes make us who we are*, TED [video] (February 2014), available at https://www.ted.com/talks/rob_knight_how_our_microbes_make_us_who_we_are?language=en.
[23] Dan Knights, Tonya L. Ward, Christopher E. McKinlay, Hannah Miller, Antonio Gonzalez, Daniel McDonald, and Rob Knight, *Rethinking "enterotypes"*, Cell Host & Microbe **16** (2014), no. 4, 433–437.
[24] Ron Kohavi, *A study of cross-validation and bootstrap for accuracy estimation and model selection*, Proceedings of the 14th international joint conference on artificial intelligence - volume 2, 1995, pp. 1137–1143.
[25] Gina Kolata, *In good health? thank your 100 trillion bacteria*, The New York Times (June 13 2012), available at http://www.nytimes.com/2012/06/14/health/human-microbiome-project-decodes-our-100-trillion-good-bacteria.html.
[26] Roie Levy and Elhanan Borenstein, *Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules*, Proceedings of the National Academy of Sciences **110** (2013), no. 31, 12804–12809, available at http://www.pnas.org/content/110/31/12804.full.pdf+html.
[27] Emeran A Mayer, Rob Knight, Sarkis K Mazmanian, John F Cryan, and Kirsten Tillisch, *Gut microbes and the brain: paradigm shift in neuroscience*, The Journal of Neuroscience **34** (2014), no. 46, 15490–15496.
[28] Kristen Mueller, Caroline Asha, Elizabeth Pennisi, and Orla Smith, *The gut microbiota*, Science **336** (2012), no. 6086, 1245, available at http://www.sciencemag.org/content/336/6086/1245.full.
[29] Michael Pollan, *Some of my best friends are germs*, The New York Times Magazine (May 15, 2013), available at http://www.nytimes.com/2013/05/19/magazine/say-hello-to-the-100-trillion-bacteria-that-make-up-your-microbiome.html.
[30] George AF Seber, *Multivariate observations*, John Wiley & Sons, 1984.
[31] small-r, *Why would anyone get a fecal transplant?*, Vimeo [Video] (February 2015), available at https://vimeo.com/119526855.
[32] Richard R. Stein, Vanni Bucci, Nora C. Toussaint, Charlie G. Buffie, Gunnar Rätsch, Eric G. Pamer, Chris Sander, and João B. Xavier, *Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota*, PLoS Comput Biol **9** (2013), no. 12.
[33] Steven N Steinway, Matthew B Biggs, Thomas P Loughran Jr, Jason A Papin, and Reka Albert, *Inference of network dynamics and metabolic interactions in the gut microbiome*, PLoS Comput Biol **11** (2015), no. 6, e1004338.
[34] The Human Microbiome Project Consortium, *Structure, function and diversity of the healthy human microbiome*, Nature **486** (2012), no. 7402, 207–214.
[35] Robert Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society. Series B (Methodological) (1996), 267–288.
[36] Andrey Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Soviet math. dokl., 1963, pp. 1035–1038.
[37] Alexa Weingarden, Antonio González, Yoshiki Vázquez-Baeza, Sophie Weiss, Gregory Humphry, Donna Berg-Lyons, Dan Knights, Tatsuya Unno, Aleh Bobr, Johnthomas Kang, Alexander Khoruts, Rob Knight, and Michael J Sadowsky, *Dynamic changes in short-and long-term bacterial composition following fecal microbiota transplantation for recurrent clostridium difficile infection*, Microbiome **3** (2015), no. 1, 10.

APPENDIX A
PRINCIPLE COORDINATE ANALYSIS

The purpose of *Principle Coordinates Analysis* (PCoA) is to represent a collection of high dimensional data in a lower dimension. Assume that one has a collection of samples $X \in \mathbb{R}^{n\times p}$ where n is the total number of samples and p is the dimension of each sample. Let $X_i \in \mathbb{R}^{1\times p}$, $i = 1, 2, \ldots, n$, be the row vectors of X as defined below $X = \begin{bmatrix} X_1^\mathsf{T} & X_2^\mathsf{T} & \cdots & X_n^\mathsf{T} \end{bmatrix}^\mathsf{T}$. The question answered in this section is how one obtains a $Y \in \mathbb{R}^{n\times k}$, $k \leq n$ with $Y_i \in \mathbb{R}^{1\times k}$, $i = 1, 2, \ldots, n$ defined as follows $Y = \begin{bmatrix} Y_1^\mathsf{T} & Y_2^\mathsf{T} & \cdots & Y_n^\mathsf{T} \end{bmatrix}^\mathsf{T}$ that is a faithful representation of X. We begin by defining the dissimilarity between samples i and j as $d(X_i, X_j)$. Then the goal of this method is to find Y such that $d(Y_i, Y_j)$ is similar to $d(X_i, X_j)$ for the dissimilarity measure of interest. Let D be a matrix composed of the sample dissimilarities where the i, j element is defined as $[D]_{ij} = -\frac{1}{2}d(Y_i, Y_j)^2$ and $B = \left(I_{n\times n} - n^{-1}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\right) D \left(I_{n\times n} - n^{-1}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\right)$, where $\mathbf{1}_n$ is an n-dimensional column vector with each entry equal to 1. The $n \times k$ dimensional representation of the $n \times p$ sample data is then $Y = \left[q_1\sqrt{\lambda_1}, q_2\sqrt{\lambda_2}, \ldots, q_k\sqrt{\lambda_k}\right]$, where $B = Q\Lambda Q^{-1}$ is the eigenvalue decomposition with eigenvalues $\lambda_i \in \mathbb{R}$, and normalized eigenvectors $q_i \in \mathbb{R}^n$ for $i = 1, 2, \ldots, n$ with $Q = [q_1, q_2, \ldots, q_n]$ and $[\Lambda]_{ii} = \lambda_i$ the diagonal eigenvalue matrix [30, Chapter 5]. Due to the fact that B is symmetric all eigenvalues and eigenvectors will be real valued. It is furthermore assumed that the eigenvalues are arranged such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

APPENDIX B
DIAGONAL STABILITY AND GENERALIZED LOTKA-VOLTERRA DYNAMICS

Definition 1. If there exists a diagonal positive matrix P such that $A^\mathsf{T}P + PA \prec 0$ then A is said to be *Diagonally Stable*.

Theorem 1 ([18, Theorem 1])**.** *If the system in* (1) *is such that the matrix A is diagonally stable and the steady state x^* is in the positive orthant ($x^* \in \mathbb{R}^n_{>0}$), then the steady state x^* is uniformly asymptotically stable for all initial conditions $x(t_0)$ in the positive orthant.*

Proof. Let $V(x,t) = 2\sum_{i=1}^n p_i(x_i - x_i^* - x_i^*\log(x_i/x_i^*))$ be the Lyapunov candidate where p_i is the i-th diagonal element of a diagonal positive matrix P such that $A^\mathsf{T}P + PA \prec 0$. Differentiating the Lyapunov candidate it follows that

$$\begin{aligned}
\dot{V}(x,t) &= 2\sum_{i=1}^n p_i\left(\dot{x}_i - x_i^*\frac{\dot{x}_i}{x_i}\right) \\
&= 2\sum_{i=1}^n p_i(x_i - x_i^*)\frac{\dot{x}_i}{x_i} \\
&= 2\sum_{i=1}^n p_i(x_i - x_i^*)\left(b_i + \sum_{j=1}^n a_{ij}x_j\right) \\
&= 2\sum_{i=1}^n p_i(x_i - x_i^*)\sum_{j=1}^n a_{ij}(x_j - x_j^*) \\
&= (x - x^*)^\mathsf{T}(A^\mathsf{T}P + PA)(x - x^*).
\end{aligned}$$

Thus the Lyapunov candidate is positive definite in $x - x^*$ and its derivative is negative definite in $x - x^*$. □