Quantifying Spatiotemporal Dynamics of Gut Microbiota and Metabolic Limitations of Cancer Cell Growth

Brian Ji

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019
ABSTRACT

Quantifying Spatiotemporal Dynamics of Gut Microbiota and Metabolic Limitations of Cancer Cell Growth

Brian Ji

In this dissertation, we develop and apply top-down, quantitative approaches to gain novel insights into various complex biological systems. Beginning at the multicellular level, we study human gut microbiome dynamics from an ecological perspective. We develop computational frameworks to enable a global understanding of the spatiotemporal variability of gut bacterial abundances. We demonstrate the utility of our frameworks to elucidate the ecological processes governing abundance changes of gut microbiota. We then shift our focus to the intracellular level by investigating the metabolic limitations of cancer cell growth. We use coarse-grained mathematical modeling to identify a major growth limitation of cancer cells associated with electron acceptor deficiency, which we then experimentally validate. Collectively, these set of approaches help to decipher the organizing principles of complex biological systems at both the individual and multicellular levels.
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................ iv

CHAPTER 1: INTRODUCTION TO THE HUMAN GUT MICROBIOME ......................... 1
   Experimental tools to characterize the gut bacterial community .................................. 3
   Factors influencing gut microbiota abundances ............................................................ 8
   Temporal dynamics of gut microbiota ........................................................................ 12
   Towards a mechanistic understanding of gut microbiota dynamics ............................ 13
   References .................................................................................................................. 16

CHAPTER 2: INTRODUCTION TO CANCER CELL METABOLISM ............................... 25
   Glucose metabolism, ATP production, glycolytic intermediates .................................. 25
   The diverse roles of glutamine in cancer cellular metabolism ....................................... 29
   Aspartate limitation in conditions of electron acceptor deficiency ............................... 32
   Lipid acquisition strategies of cancer cells .................................................................. 33
   References .................................................................................................................. 36

CHAPTER 3: QUANTIFYING SPATIOTEMPORAL DYNAMICS AND NOISE IN
ABSOLUTE MICROBIOTA ABUNDANCES USING REPLICATE SAMPLING ............... 41
   Author statement ........................................................................................................ 41
   Abstract ...................................................................................................................... 42
   Results ....................................................................................................................... 42
   Conclusion .................................................................................................................. 53
   Methods ...................................................................................................................... 55
   References .................................................................................................................. 68
CHAPTER 4: MACROECOLOGICAL RELATIONSHIPS IN THE DYNAMICS OF GUT MICROBIOTA

Author statement ............................................................................................................................................. 71
Abstract ............................................................................................................................................................. 72
Results ................................................................................................................................................................. 73
Conclusion .......................................................................................................................................................... 85
Methods ............................................................................................................................................................. 87
References ........................................................................................................................................................... 92

CHAPTER 5: LIPIDS ARE A MAJOR GROWTH LIMITATION IN CELLS DEFICIENT FOR ELECTRON ACCEPTORS

Author statement ............................................................................................................................................. 97
Abstract ............................................................................................................................................................. 98
Results ................................................................................................................................................................. 99
Discussion ........................................................................................................................................................ 107
Computational Methods ............................................................................................................................... 110
References ........................................................................................................................................................ 114

CHAPTER 6: CONCLUSION ............................................................................................................................... 116

Summary .......................................................................................................................................................... 116

DIVERS: A generalized dual-reporter method for sequencing studies ......................................................... 118
Towards the investigation of human gut microbiota dynamics across populations ........................................ 119
Coarse-grained approaches for studying cancer cell physiology ............................................................... 121
References ........................................................................................................................................................ 123

APPENDIX A: SUPPLEMENTARY FIGURES AND TABLES FOR CHAPTER 3 ............... 126
APPENDIX B: SUPPLEMENTARY FIGURES AND TABLES FOR CHAPTER 4 ........... 145
Author statement.................................................................................................................. 145

APPENDIX C: FULL DESCRIPTION OF VARIANCE AND COVARIANCE
DECOMPOSITION MODELS FOR CHAPTER 3 ........................................................................ 154
Author statement.................................................................................................................. 154
Variance decomposition model.............................................................................................. 155
Covariance decomposition model.......................................................................................... 165
References................................................................................................................................ 171

APPENDIX D: SUPPLEMENTARY FIGURES AND TABLES FOR CHAPTER 5 ........... 172
Author statement.................................................................................................................. 172
LIST OF FIGURES

Figure 3. 1. Variance decomposition of gut bacterial abundances using DIVERS. ......................... 47
Figure 3. 2. Temporal and spatial variation of human gut microbiota. ........................................ 49
Figure 3. 3. Decomposition of pairwise OTU abundance correlations in the human gut microbiome. .................................................................................................................................................. 50
Figure 3. 4. Variance decomposition applied to the dynamics of soil bacterial abundances. ...... 53

Figure 4. 1. Daily changes in the abundances of gut microbiota.................................................. 76
Figure 4. 2. Long-term stability of gut microbiota abundances.................................................... 78
Figure 4. 3. Residence and return times of gut microbiota............................................................ 79
Figure 4. 4. Taylor’s power law in the gut microbiome. ............................................................... 81
Figure 4. 5. Dynamics of gut microbiota in mice fed different diets........................................... 84

Figure 5. 1. Lipids biosynthesis represents a major demand for NAD+ and confers sensitivity of cancer cells to mitochondrial inhibition.................................................................................. 100
Figure 5. 2. Pyruvate the rescues growth sensitivity of cells to ETC inhibition by supporting de novo lipogenesis................................................................................................................. 104
Figure 5. 3. Lipids are the primary metabolic limitation of cancer cells limited for electron acceptors. ................................................................................................................................. 106

Figure A 1. Validation of the spike-in sequencing approach to estimate total bacterial loads in collected fecal and soil samples........................................................................................................ 128
Figure A 2. Variance decomposition of OTU relative abundances in the human gut microbiome. ................................. 129

Figure A 3. Variance decomposition for high and low-abundance OTUs in the human gut microbiome. ..................... 130

Figure A 4. Pairwise OTU abundance correlations in the human gut microbiome. .................................................. 131

Figure A 5. Correlations of OTU abundances within and between different phyla in the human gut. ................................. 132

Figure A 6. Average correlations of OTU abundances within and between different microbial families in the human gut. ........................................................................................................ 133

Figure A 7. Pairwise abundance correlations of gut bacterial OTU abundances within phyla at different phylogenetic distances. ........................................................................................................ 134

Figure A 8. Sampling sites from which soil bacteria were collected for 16S rRNA sequencing. ................................. 135

Figure A 9. Application of DIVERS to a Central Park soil community. ................................................................. 136

Figure A 10. Decomposition of pairwise OTU abundance correlations in the soil microbiome. .................................. 137

Figure B 1. Daily growth rate distributions of human and mouse gut microbiota. .................................................. 146

Figure B 2. Growth rate distributions in diverse ecological communities and economic systems. ................................. 147

Figure B 3. Long-term dynamics of gut microbiota abundances in humans and mice. ............................... 148

Figure B 4. Long-term behavior of individual OTU abundances. ................................................................. 149

Figure B 5. Residence and return times of gut microbiota in humans and mice. ..................................................... 150

Figure B 6. Taylor’s power law in the human and mouse gut microbiome. ......................................................... 151
Figure C 1. Statistical model for the decomposition of bacterial abundance variances in the fecal microbiome. ........................................................................................................................................ 157

Figure C 2. Decomposing the covariance in pairs of bacterial species abundances. .................. 166

Figure D 1. Lipid species concentrations in serum-delpidated media........................................... 173
Figure D 2. Lipid synthesis rates in lipid-depleted media conditions............................................. 174
Figure D 3. Oxidative lipid synthesis in lipid-replete or lipid-depleted conditions....................... 175
I am grateful to countless individuals who have helped to guide me through my scientific journey. To my undergraduate research mentors Dr. David Hsu and Cheng Guan Kuoy at UW-Madison, thank you for teaching me to pursue questions that first and foremost capture my imagination.

To my PhD advisor, Dennis Vitkup, thank you for showing me over the years what it means to be a true investigator of science. You have shaped my scientific outlook and perspective, and introduced me to areas in science I would have otherwise never been exposed to - for that I’ll always be grateful.

To my lab mate and friend Purushottam Dixit, thank you for tolerating all my questions and always having the answers to them when I didn’t have them myself.

To all the past and present members of the Vitkup lab: Andy Chiang, Jonathan Chang, German Plata, Eugenia Lyashenko, Dinara Usmanova, Mariam Konate, Jie Hu and Kostya Tchourine, thank you for creating a supportive atmosphere. While we all worked on different things, I’ve learned from every one of you.

To my collaborators at the Vander Heiden lab at MIT: Zhaoqi Li, Lucas Sullivan, Aaron Hosios and Matt Vander Heiden, thank you for sharing your knowledge with me. I can’t overstate how influential our discussions were to my scientific growth.
To my collaborators at the Wang lab: Ravi Sheth and Harris Wang, thank you for sharing your honest scientific perspectives. A second thank you to Ravi for helping me bring to life what we initially thought was a crazy idea.

To my all my former and current medical school and MD/PhD classmates, thank you for being a source of support. Special thank you to my roommate Danny Echelman who allowed me to store “samples” in the freezer.

To my parents, thank you for your love, and for being examples of earnestness and hard work.

To my brother Andrew, thank you for being a role model to me.

And to my fiancée Amanda, thank you for your daily love and support.
CHAPTER 1: INTRODUCTION TO THE HUMAN GUT MICROBIOME

The set of microorganisms residing within the human gastrointestinal tract, collectively referred to as the human gut microbiota, has seen a recent explosion of research interest over the past two decades (Arumugam et al., 2011; Huttenhower et al., 2012; Lloyd-Price et al., 2017; Qin et al., 2010; Yatsunenko et al., 2012). Bacterial cells in the human gut microbiome are similar in number to those of the host, and contain a roughly two orders of magnitude increase in gene content relative to the human genome (Qin et al., 2010). The symbiosis between humans and their gut bacteria likely extends back thousands of years and has been continuously shaped by recent human evolution (Walter and Ley, 2011). It is now becoming evident that the complex community of microbes residing within our guts plays an integral role in human health and disease (Cho and Blaser, 2012; Clemente et al., 2012). One of the very first connections between these microorganisms and host physiology was established in seminal studies demonstrating that germ-free mice lacking a gut microbiome were resistant to diet-induced weight gain and glucose intolerance compared to their colonized counterparts (Backhed et al., 2004; Bäckhed et al., 2007). Numerous work has since established that the gut microbiome performs a diverse array of functions, including but not limited to the breakdown of ingested dietary nutrients (Sonnenburg and Bäckhed, 2016; Sonnenburg and Sonnenburg, 2014), the modulation of the human immune system (Arpaia et al., 2013; Furusawa et al., 2013; Smith et al., 2013), the prevention of pathogenic infection (Ng et al., 2013; Thaiss et al., 2018) and the maintenance of metabolic homeostasis associated with the host circadian clock (Thaiss et al., 2014, 2016).
Given the enormous taxonomic and physiological diversity of bacteria in any individual gut, in concert with a highly dynamic and continuously changing environment, the human gut microbiome exhibits all the hallmarks of a complex ecosystem (Costello et al., 2012). As such, ideas from theoretical ecology have been increasing applied to understand the primary forces and mechanisms shaping bacterial abundances in the gut (Costello et al., 2012). Gut bacteria participate in multimodal and complex interactions with other species (Coyte et al., 2015a), the host epithelium (Bäckhed et al., 2005), and the host immune system (Lozupone, 2018). They are also subjected to daily stochastic and environmental perturbations, including the intake of dietary macronutrients and medications by the host (David et al., 2014a; Maier et al., 2018). In contrast to other ecosystems however, the collective state of the human gut microbiome is intimately connected to the health status of the host, and this connection may be further entrenched by numerous host-bacteria feedback interactions (Foster et al., 2017). Crucially, a growing number of studies have now associated alterations to gut microbiota composition, diversity and abundances, to various human diseases ranging from autoimmune disorders (Frank et al., 2007a; Maeda and Takeda, 2017) and metabolic syndrome (Le Chatelier et al., 2013; Ley et al., 2005; Qin et al., 2010; Thaiss et al., 2018; Turnbaugh et al., 2006; Wang et al., 2012) to cancer (Kostic et al., 2012). While the mechanisms underlying many of these associations remain to be elucidated, increasing evidence suggests that the microbiome may play a causal role in at least several of these diseases (Ridaura et al., 2013; Smith, M.I., Yatsunenko et al., 2013; Turnbaugh et al., 2006). Because of its relative malleability, targeting the gut microbiome represents an exciting avenue for novel and personalized treatment modalities (Sonnenburg and Bäckhed, 2016; Zmora et al., 2016). However, as with any ecosystem, understanding the response of the gut microbiome to targeted interventions first requires an understanding of the primary forces that shape bacterial abundances in the healthy
human gut. Numerous research efforts have begun to build on this understanding through the systematic characterization of gut bacteria across large population cohorts, as well as high-resolution profiling of human gut microbiota abundances across space and time within single individuals.

**Experimental tools to characterize the gut bacterial community**

Rapid advances in next-generation sequencing technologies have changed the landscape of microbial ecology and studies of the gut microbiome (Caporaso et al., 2011a; Gohl et al., 2016; Kozich et al., 2013; Quince et al., 2017). These methodologies have enabled quantitative characterization of the vast taxonomic diversity of gut microbiota (Huttenhower et al., 2012; Lloyd-Price et al., 2017). Furthermore, unbiased profiling of microbial genomes has begun to identify the functional capacities and repertoires of entire gut microbiome communities (Huttenhower et al., 2012; Qin et al., 2010). As sequencing efforts have become cheaper, technologies may now be leveraged to understand patterns of microbiome variation across large populations of healthy and diseased cohorts (Falony et al., 2016; Rothschild et al., 2018; Zhernakova et al., 2016). Furthermore, the temporal dynamics of gut microbiota within multiple individual people can now be monitored at unprecedented temporal resolution (Caporaso et al., 2011b; David et al., 2014b; Flores et al., 2014). However, as with any rapidly developing technologies, quantitative frameworks to harness the vast amounts of generated data are required (Gilbert and Dupont, 2011).

*16S rRNA sequencing*
Marker gene-based surveys of microbial communities have proven to be the most simple and cost-effective method of quantifying bacterial abundances in the gut (Pollock, 2018). Marker gene surveys refer to the isolation and characterization of specific genes in the community, which serve as proxies for the bacteria containing those genes. In order for these studies to be effective, the gene in question must meet a set of criteria. First, it must be present in all members of the gut community. Second, the gene sequence must be similar for bacteria with similar genome content and differ in phylogenetically distinct taxa. Finally, the dissimilarity between gene sequences from two different bacteria should ideally reflect the evolutionary distance between the bacteria (Morgan and Huttenhower, 2012). Fortunately, the field of microbial ecology has identified the 16S rRNA gene, a member of the small bacterial ribosomal subunit, as an ideal marker gene candidate. As the 16S rRNA gene is relatively easy to target, isolate and sequence, the challenge of characterizing bacterial taxonomic diversity in the gut has become a matter of quantifying the relative copy numbers of 16S gene sequence variants in the community (Janda and Abbot, 2007).

In a typical 16S rRNA marker gene study, bacterial DNA is first extracted from a collected fecal sample. While the exact degree to which the fecal bacterial community reflects the in vivo gut community remains actively investigated (Zmora et al., 2018), relative abundances in the fecal microbiome remain a simple and practical proxy for gut bacteria. Targeted primers to highly conserved regions of the 16S gene flanking variables sub-regions are utilized to PCR amplify the region of interest. The resulting amplicons are subjected to next-generation sequencing, resulting in set of short reads representing the sequence variants of the 16S gene in the original fecal community (Caporaso et al., 2011a; Kozich et al., 2013).
To extract information on relative taxa abundances from short reads, 16S sequence variants are typically clustered by sequence similarity. Clustering may be done by mapping reads to a reference database of known 16S rRNA sequences (DeSantis et al., 2006; Quast et al., 2013), or may performed de novo. Mapping reads to a reference database may filter out spurious sequences that reflect chimeras or sequencing errors, but may inadvertently discard any novel taxonomic diversity (Morgan and Huttenhower, 2012). The resulting clusters are referred to as operational taxonomic units (OTUs) and represent sets of similar bacteria up to a specified sequence similarity threshold. The 97% sequence identify threshold in 16S rRNA is commonly taken to be a designation of a single bacterial species. Taxonomies may be assigned to different OTU sequences based on classifiers trained on large database sets (Wang et al., 2007). The resulting OTU abundances provide valuable quantitative information of taxonomic diversity within and between different gut microbial communities. These abundances may then be investigated across different host states to understand disease pathophysiology (Frank et al., 2007a; Ley et al., 2005; Segata et al., 2011), in conjunction with network inference techniques to predict microbe-microbe interactions (Faust and Raes, 2012; Faust et al., 2012; Zhang et al., 2014) and over time to decipher microbiome dynamics (Bucci and Xavier, 2014; Faust et al., 2015; Stein et al., 2013).

Whole-metagenome shotgun sequencing

16S rRNA amplicon sequencing studies have gained enormous popularity because of their cost-effectiveness and relative ease of sample preparation and sequencing. However, 16S rRNA studies have a number of important limitations. First, they are sensitive to PCR amplification bias (Gohl
et al., 2016). Second, copy numbers of the 16S rRNA gene may vary between bacterial taxa (Langille et al., 2013). In this way, the relative abundances of 16S gene copies may not reflect the cellular composition of a community. Finally, 16S gene sequences cannot resolve high-resolution taxonomic differences. Indeed, large divergence of bacterial phenotypes has been observed for bacteria with nearly identical 16S gene sequences (Plata et al., 2015).

Many of the aforementioned limitations of 16S rRNA sequencing may be resolved with the use of whole-metagenome shotgun sequencing (WMGS) (Quince et al., 2017). Metagenomic shotgun sequencing refers to the unbiased sequencing of the entire genomic content in a microbial community, across multiple different bacterial taxa (in contrast to the targeted, marker-based approach). WMGS enables investigation of microbial community composition at the individual species and strain level. Furthermore, estimated abundances are less sensitive to PCR amplification biases. With costs of sequencing continuing to decrease, it is likely that WMGS will play a larger role in gut microbiome studies, where it has previously been impractical or infeasible (Quince et al., 2017). Indeed, a recent study of shallow shotgun sequencing demonstrated that this methodology captured much of the information content of 16S rRNA sequencing at a comparable cost per sample. (Hillmann et al., 2018) However, WMGS is not without its own challenges, particularly in the bioinformatic limitations associated with taxonomic profiling and abundance estimation.

Two major classes of methods exist for taxonomic profiling of gut microbial communities based on WMGS. The first is the de novo assembly of metagenomes and binning of assembled contigs for taxonomic classification (Quince et al., 2017). The second are assembly-free methods that
involve mapping of reads directly to reference genome databases. Here, we discuss methods based on the latter approach.

A primary challenge of assembly-free metagenomic taxonomic profiling is the requirement of existing, fully annotated and characterized genomes. These reference genomes must fully span the original taxonomic diversity that exists in the gut, which is unknown a priori. However, with thousands of newly sequenced genomes produced each year, the problem of insufficient reference genomes will likely be mitigated, especially in the human gut microbiome for which the diversity of reference genomes is already large. A second challenge is the computational assignment of short reads to reference sequence genomes. Mapping of short reads by brute force sequence alignment is highly inefficient and results in numerous false positive assignments (Brady and Salzberg, 2011, 2009; Peabody et al., 2015). A suite tools have been recently developed to address the computational challenge of taxonomic abundance estimation from WMGS. One popular method, MetaPhlAn (Segata et al., 2012), builds a precomputed database of clade-specific marker genes, to which short reads are subsequently mapped. Because these marker genes represent a small fraction of the overall genomic content, the computational time of MetaPhlAn is much smaller relative to naïve approaches. Another tool, Kraken (Wood and Salzberg, 2014), performs taxonomic read classification by matching read k-mers to a large, precomputed k-mer database based on a set of reference genomes. These and other bioinformatics tools hold promise for the use of WMGS in characterizing not only the gut microbiome, but microbial communities across other highly diverse environments.

*Limitations associated with relative abundances*
An inherent limitation of both 16S rRNA sequencing and WMGS is that taxonomic abundances are reported as relative fractions of the entire community. This effect, known as compositionality, is problematic because information on absolute abundance changes in the bacterial community are masked (Vandeputte et al., 2017). That is, if the total bacterial loads across different gut communities are highly variable, changes in the relative abundances of different bacteria in the community do not reflect changes in their true absolute abundances. Compositionality can lead to a number of artifacts in downstream data analysis, including artificial correlations between taxa pairs (Friedman and Alm, 2012) and the misidentification of taxa associated with disease states (Vandeputte et al., 2017). A number of computational approaches have been developed to correct for artifacts of compositionality that involve data transformations on original relative abundances (Friedman and Alm, 2012; Silverman et al., 2017). Other methods have attempted to directly measured the total number of bacteria in each sample, enabling the estimation of absolute abundances of individual taxa (Stämmler et al., 2016; Tkacz et al., 2018; Vandeputte et al., 2017). These and other studies have highlighted the limitations of relative abundance profiling and argue for the necessity of absolute abundance measurements in future studies of the gut microbiome.

**Factors influencing gut microbiota abundances**

*Host-intrinsic sources of variability*

Studies of the gut microbiome across large population cohorts been instrumental in identifying the factors that contribute to microbiota figurations in humans and how they may be perturbed in
disease. These factors may be broadly classified as intrinsic or extrinsic to the host, roughly corresponding to ecological mechanisms occurring within individual hosts as opposed to outside environmental forces. One of the intrinsic factors receiving much attention in the human gut microbiome field is the notion of enterotypes, or stable configurations of microbiota observed across human populations, independent of age or geography (Arumugam et al., 2011). Different enterotypes, broadly classified by their predominating taxa, are believed to represent inherently stable configurations of gut bacterial communities, though their existence has been called into question recently (Knights et al., 2014). Host genetic has also been proposed to play a significant role in shaping gut communities across different individuals (Bonder et al., 2016; Goodrich et al., 2014; Wang et al., 2016). Interestingly, twin studies have identified a number of highly heritable taxa, including the Christensenellaceae, a family that was shown to play a causative role in weight-gain prevention in mouse models (Goodrich et al., 2014). However, a more recent population wide analysis reported that host genetics plays a minor role in shaping inter-individual gut microbiota differences, estimating that genetic factors may explain only ~2% of the variance in gut microbiota abundances (Rothschild et al., 2018).

*Host-extrinsic sources of variability*

Environmental factors are known to the shape species abundances of various ecosystems and the gut microbiome is no exception. Initial events in early childhood are believed to have long-lasting impacts on the gut microbiome later in adulthood, including the very first maternal colonization events associated with either vaginal birth delivery or Caesarian section (Tamburini et al., 2016). Taxonomic composition continues to change throughout childhood, influenced by factors such as
breast or formula feeding, with further increases in taxonomic diversity into adulthood (Stewart et al., 2018; Vatanen et al., 2018; Yatsunenko et al., 2012).

In adults, population-wide studies have identified a number of co-variates explaining some of the inter-individual variability in gut microbiome configurations. These factors include, age, blood parameters, stool transit time and consistency, dietary intake, and medication use, and explain roughly 20% of inter-individual gut microbiome variability (Falony et al., 2016; Rothschild et al., 2018). Interestingly, a recent study reported that a large percentage of the over 1,000 human-targeting marketed drugs that were investigated had an adverse impact on at least one gut commensal (Maier et al., 2018), suggesting medication use should be properly accounted for when conducting metagenome-wide associations studies in large population cohorts (Forslund et al., 2015).

Dietary effects of gut microbiota

Of all the factors that are known to influence gut-associated bacteria, diet has perhaps received the most attention due to the central role of diet-microbiota interactions in modulating human health (Sonnenburg and Bäckhed, 2016). A growing consensus established from mouse models is that diet can rapidly and reproducibly alter the taxonomic composition and abundances of gut microbiota, within the time scale of single day (Carmody et al., 2015; David et al., 2014a). Using a diet-induced mouse model of obesity, Turnbaugh and colleagues first demonstrated that a diet high in fat and simple sugars (HFHS diet) could significantly alter gut microbiota of mice compared to littermates fed a low-fat plant polysaccharide-rich control (LFPP) diet. In mice, this
effect has since been shown to be independent of host genetics or the immune system, suggesting that gut microbiota respond directly to luminal concentrations of dietary nutrients (Carmody et al., 2015). Strikingly, the change in microbiota composition in the HFHS mice was associated with increase adiposity and weight gain, a phenotype that was transferrable when the gut microbiomes of the HFHS mice were inoculated into the guts of germ-free mice (Turnbaugh et al., 2006). This demonstration of a causal association between an altered microbiota and disease phenotype raises the possibility that host diet be used as modulator of microbiota configurations, with the potential for beneficial outcomes to host health (Sonnenburg and Bäckhed, 2016).

Dietary changes may also significantly alter the gut microbiota of humans. Shifts from baseline diets to plant-based or an animal protein-based diets for several days caused rapid blooms of certain taxa, as well as significant changes at the meta-transcriptomic level (David et al., 2014a). However, in contrast to mice, short-term dietary perturbations in humans do not alter the inter-individual differences between human gut microbiomes (David et al., 2014a; Wu et al., 2011). The relatively stable microbiota signatures in humans raises the possibility that our gut microbiomes have experienced irreversible changes over time, driven in part by a Westernized diet and lifestyle (Sonnenburg and Sonnenburg, 2014). To explore this idea, Sonnenburg et. al. demonstrated in a mouse model that a diet low in plant-based carbohydrates, taken to reflect a “Westernized” diet, resulted in significant and irreversible loss of taxonomic diversity over multiple generations. The loss of taxonomic diversity correlated with the loss of specific gene and metabolic enzymes required to break down dietary fiber. Interestingly, Western populations in the United States and Europe have significantly reduced taxa diversity compared to more rural, agrarian societies (Yatsunenko et al., 2012). This raises the possibility that the high incidence of
Western-associated diseases, such as obesity and the metabolic syndrome, may in part be a result of the loss of bacteria in Western populations that otherwise decrease the risk of such diseases in other populations.

**Temporal dynamics of gut microbiota**

In contrast to static snapshots across entire populations, the temporal dynamics of ecosystems contain information on the underlying biological processes operating within individual communities. A number of recent longitudinal studies have begun to paint a picture of the human gut microbiome as a highly complex and dynamic ecosystem (Caporaso et al., 2011b; David et al., 2014b; Flores et al., 2014). Despite large abundance fluctuations from one day to the next, the composition of the gut community is in fact relatively stable. That is, each individual appears to harbor a relatively unique set of strains that once present, may reside in the gut of that individual for many years or even decades (Faith et al., 2013). The long-term stability of bacteria, in part driven by negative interactions between gut community members (Coyte et al., 2015a), is often taken to reflect a healthy community. Related to long-term stability, previous longitudinal studies have demonstrated that the gut microbiome is also relatively resilient to perturbations (Relman, 2012). For example, upon exposure to antibiotic treatment, individual gut microbiomes exhibit recovery to the pre-treated state (Suez et al., 2018), though the rate and degree of recovery appears to be person-specific (Dethlefsen and Relman, 2011).

Two seminal studies have conducted the longest and highest resolution time series profiling of human gut microbiota to date. In the study of David and Alm, the fecal microbiomes of two healthy
male individuals were tracked daily over the course of several hundred days (David et al., 2014b). Analysis of their gut microbiomes revealed that for long stretches of time, the composition of gut microbiota remained relatively stable. Only after a series of perturbation events to the guts of each individual did their microbiome configurations become dramatically altered. One individual, who traveled to a third-world country for several weeks, experienced a surge in Proteobacteria abundances during the travel period. Interestingly, the microbiota of this individual become more similar to pre-travel conditions upon return to the United States. The gut microbiome of the second individual, after experiencing an enteric infection, remained in an altered state for the duration of the study. This suggests that different ecological perturbations may be associated with reversible or irreversible changes, and that the gut microbiome may encode memory of these disruptions. In a second study, two individuals were tracked at daily resolution for several hundred days at multiple body sites, including the gut, skin and tongue (Caporaso et al., 2011b). Interestingly, the primary source of variability between microbiomes was body site, which was much greater than differences between individuals at the same site. These inter-site differences remained highly stable over the duration of the study, suggesting that long-term stability may be a property of multiple human microbiome communities.

Towards a mechanistic understanding of gut microbiota dynamics

The temporal behavior of gut microbiota involve highly complex interactions with other bacteria, the host epithelium and immune system, and the surrounding environment. A key challenge in understanding the dynamics of gut microbiota is the development of tractable, quantitative approaches that harness information from experimental measurements to enable mechanistic
insights into dynamical behavior. One popular class of methods have modeled gut bacterial dynamics following a generalized Lotka-Volterra (GLV) model, given by the following set of equations (Bucci and Xavier, 2014; Buffie et al., 2015; Stein et al., 2013):

\[
\frac{dx_i}{dt} = r_i x_i \left(1 - \frac{x_i + \sum_{j=1}^{n} A_{ij} x_j}{K_i}\right) \tag{1.1}
\]

Here, \(x_i\), \(r_i\) and \(K_i\) correspond to the abundance, intrinsic growth rate, and carrying capacity of taxa \(i\) respectively, and \(A_{ij}\) is a matrix encoding pairwise interactions between different members of the community. Given the measured time series trajectories \(x_i(t)\) of each individual taxa \(i\) in the community, equation (1) can be used to learn the underlying parameters in the system. These include the intrinsic growth rates of taxa as well as any synergistic or antagonistic interactions between pairs of taxa in the community. This approach has been applied successfully to identify both probiotic and prebiotic treatment candidates in a murine model of \(C.\) difficile colitis (Buffie et al., 2015). However, a major limitation of such modeling approaches is the issue of parameter indeterminacy, which refers to the degeneracy of parameter sets that are equally compatible with observed data. If the number of unknown parameters vastly exceeds the number of experimental measurements, which is common in the setting of the gut microbiome, parameter indeterminacy renders the parameter inference process highly challenging.

To bypass this issue, other studies have used the GLV framework to understand broader-level questions, such as how general classes of interactions (cooperative, antagonistic, and symbiotic) contribute to the overall stability of the gut microbiome as well as other ecosystems (Allesina and Tang, 2012; Coyte et al., 2015a). Interestingly, cooperative behavior of gut bacteria was predicted
to reduce the overall stability of dynamics, while mutually antagonistic interactions, such as those induced by competition for finite resources, increased the long-term stability of ecosystems (Coyte et al., 2015b). Another class of studies have either abandoned parameterization altogether or focused on simplified and experimentally tractable model microbial ecosystems to build a bottom-up understand of microbial dynamics and interactions (Celiker and Gore, 2014; Friedman et al., 2017; Peng et al., 2018; Vega and Gore, 2017). For example, outcomes of pairwise bacterial competitions have been used to predict the outcomes of larger community members of the same bacteria, though the accuracy of this approach declines as the size of the overall community grows (Friedman et al., 2017).

In contrast to previous mechanistic or bottom-up approaches, top-down statistical approaches have been successfully applied to understand the dynamics of a number of diverse macroscopic ecosystems (Azaele et al., 2016; Keitt and Stanley, 1998; Marquet et al., 2005; Sun et al., 2015). These approaches focus on quantifying global aspects of species dynamics through a series of statistical relationships. Notably, while such macroecological analyses have been applied to understand static patterns of bacterial species abundances (Li and Ma, 2016; Shoemaker et al., 2017), a comprehensive analysis of gut microbiota dynamics has not been conducted. Given the complexity of gut bacterial dynamics across time and space, top-down approaches may provide tractable phenomenological descriptions, elucidating patterns of microbiota dynamics in the human gut that may not otherwise be captured. This sets the stage for the development and application of novel, quantitative frameworks that leverage high-resolution time series profiling of the human gut microbiome to unravel the spatiotemporal dynamics of gut bacteria.
References


Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A.,


Cancer cells must alter their metabolic processes to support rapid proliferation (Cairns et al., 2011; Cantor and Sabatini, 2012; Ward and Thompson, 2012a). Because of its fundamental connection with tumor survival and growth, metabolism of cancer cells has been receiving renewed attention nearly a century after the initial, seminal findings of Otto Warburg, who first described altered patterns of glucose metabolism in proliferating ascites tumor cells (Warburg, 1925, 1956). Now, with the rapid emergence of high-throughput technologies, recent genetics, genomics, transcriptomics, metabolomics, fluxomics, and computational studies have offered new insights into specific metabolic alterations in cancer cell lines and tumors (Hu et al., 2013; Jain et al., 2012; Peng et al., 2018). Genetic and functional approaches have also revealed that many essential metabolic processes are under direct regulatory control of known oncogenes and tumor suppressors (Ward and Thompson, 2012b). The accumulated knowledge about the role of metabolic changes in tumorigenesis is now being applied to develop anti-cancer therapeutics that target specific metabolic enzymes in various disease contexts (Vander Heiden, 2011).

**Glucose metabolism, ATP production, glycolytic intermediates**

Perhaps the most striking metabolic alteration associated with cancer cells and tumors is an elevated rate of glucose consumption and associated efflux of lactate. In contrast to normal cells, rapidly proliferating cancer cells usually convert the majority of imported glucose to lactate, even in the presence of sufficient oxygen to support oxidative phosphorylation, a phenomenon known as aerobic glycolysis or the Warburg effect (Vander Heiden et al., 2009; Warburg, 1925). Although
the role of oxidative phosphorylation has been shown to be heterogeneous across different tumor types and cell lines, the substantial increase in glucose uptake has been exploited by the widely used FDG-PET scan for imaging tumors in patients.

The preferential excretion of glucose-derived lactate despite the presence of oxygen is unexpected, given the primary routes of ATP synthesis in mammalian cells and the considerable energetic demands of rapid cell growth. Mammalian cells produce ATP through two primary routes, which correspond to the fates of glucose-derived intracellular pyruvate. When pyruvate enters the mitochondria and is oxidized into carbon dioxide in the TCA cycle, oxidative phosphorylation in the mitochondria produces roughly 36 molecules of ATP per glucose molecule (Heiden et al., 2009). On the other hand, conversion of pyruvate to lactate and its subsequent excretion from cells produces only 2 molecules of ATP per glucose molecule. On the basis of ATP yield, this represents over an order of magnitude reduction in the amount of ATP produced through the fermentation of glucose to lactate relative to complete oxidation. Thus, the ubiquity of aerobic glycolysis in cancer cells remains an open question.

Rapid rates of glucose uptake and the excretion of incompletely oxidized carbon waste products are common to many rapidly proliferating unicellular organisms (Basan et al., 2015; Varma and Palsson, 1994; Vemuri et al., 2006; Zhuang et al., 2011). In laboratory strains of E. coli, the excretion of acetate from the cell occurs at a critical growth rate, past which acetate secretion rates increase and respiration rates decrease linearly with the proliferation rate respectively (Basan et al., 2015). That is, bacterial cells actively increase glycolytic fluxes and suppress oxidative phosphorylation at rapid proliferation rates. This strategy is believed to result from optimal
allocation of proteins into different metabolic sectors to maximize the bacterial growth rate. While ATP yield per molecule of glucose from glycolysis is much less than that from oxidative phosphorylation, the ATP yield per mass of protein is surprisingly higher for the collective enzymes in the glycolytic pathway than those associated with oxidative phosphorylation (Shlomi et al., 2011). This stems from a combination of the overall fewer total number of glycolytic enzymes, their lower molecular weights, and higher enzyme catalytic constants relative to those associated with the TCA cycle and oxidative phosphorylation. Therefore, glycolysis in bacteria can maintain similar rates of ATP production using lower total concentration of enzymes in the cell, enabling increased expression of other intracellular proteins required to support rapid proliferation, such as ribosomal proteins (Scott et al., 2010). Whether a similar mechanism is true for cancer cells remains to be determined.

Another explanation for the rapid rates of glucose uptake in cancer cells is that ATP production may not be the sole limitation to cancer cell and tumor growth. Cancer cells must produce a number of collective biomass precursors, including lipids, nucleotides, amino acids and carbohydrates (Vander Heiden et al., 2009). Rapid generation of cofactors such as NADPH, which provide reducing equivalents for de novo lipid synthesis, and NAD+ to sustain various oxidative reactions are also required. Notably, products from a number pathways directly branching from glycolysis fulfill these requirements. Therefore, the rapid uptake of glucose may be important for the shunting of various glycolytic intermediates into anabolic pathways that provide both the carbon building blocks and cofactors required for growth.
The first major branch off of glycolysis is the oxidative pentose phosphate pathway (PPP), beginning with the conversion of glucose-6-phosphate (G6P) into 6-phosphogluconolactone (6PG) by the rate-limiting enzyme glucose-6-phosphate dehydrogenase (G6PD). The oxidative branch of the PPP is the primary source of cytosolic NADPH in a number of cancer cell lines (Fan et al., 2014). The generation of NADPH not only provides reducing equivalents for lipid biosynthesis, but is also an important source of antioxidants to defend cells from intracellular reactive oxygen species (ROS) accumulation. Indeed, the activity of the specific isoform of the glycolytic enzyme PKM2 has been shown to be attenuated in the presence of high intracellular ROS, thereby diverting flux into the oxidative PPP to maintain NADPH homeostasis and protect from ROS-induced cell death (Anastasiou et al., 2011). Furthermore, both the oxidative and non-oxidative branches of the PPP are important sources of ribose-5-P, which serves as the carbon sugar backbone of nucleotides. Pancreatic tumors harboring K-Ras mutations have been shown to elevate expression of enzymes in the non-oxidative PPP, which is required for rapid nucleotide production (Ying et al., 2012).

De novo synthesis of the non-essential amino acid serine represents another important branch point off of glycolysis. In cancer cells, the first and rate limiting step of this pathway is mediated by the enzyme 3-phosphoglycerate dehydrogenase (PHGDH). Many ER-negative breast cancers exhibit focal amplifications of PHGDH, and inhibition of the serine biosynthesis pathway in the context of high PHGDH expression limits tumor progression in vivo (Possemato et al., 2011). The end product of this pathway, serine, is not only important for protein synthesis, but also for fueling folate-mediated one-carbon metabolism (Locasale, 2013). This complex series of interconnected reactions transfers one carbon units donated from serine towards the synthesis of pyrimidine and
purine nucleotides, maintenance of DNA methylation and epigenetic states, and the production of NADPH. NADPH production mediated by one carbon metabolism has been recently shown to be important for protecting against high intracellular ROS associated with anchorage-independent growth and metastasis (Piskounova et al., 2015). One carbon units derived from serine have also been recently shown to be required for mitochondrial translation and proper ETC function, through the modification of certain tRNAs (Minton et al., 2018; Morscher et al., 2018). Finally, through the activity of phosphoserine amino transferase (PSAT), the serine biosynthesis pathway represents an important route for alpha-ketoglutarate production, and glutamine-mediated anaplerosis of the TCA cycle (Possemato et al., 2011).

**The diverse roles of glutamine in cancer cellular metabolism**

In addition to glucose, the non-essential amino acid glutamine, the most abundant amino acid in cell culture media and human plasma, has been long recognized as an important contributor to cancer cell growth (DeBerardinis and Cheng, 2010; DeBerardinis et al., 2007; Wise et al., 2008). Glutamine can be an important source of energy, a source of nitrogen for nucleotides and amino acids, and also a carbon source for various biosynthetic precursors.

Glutamine plays an important role in nucleotide metabolism by donating its γ-nitrogen for de novo synthesis of both purines and pyrimidines. In many proliferating cells in culture, however, the glutamine consumption rate exceeds the nucleotide biosynthetic need by almost an order of magnitude (Jain et al., 2012). The extra glutamine is used in multiple biosynthetic and energetic processes after its conversion to glutamate. The deamination of glutamine to glutamate occurs in
many cancer cells primarily through the activity of glutaminase (GLS). Notably, it was demonstrated that the oncogenic master regulator Myc may transcriptionally induce the expression of glutamine transporters SLC7A5/SLC7A1, and post-transcriptionally increase the level of GLS1 (Gao et al., 2009).

Glutamine’s α-nitrogen carried by glutamate is dispersed into pools of various nonessential amino acids. This occurs through the activity of several transaminases, particularly alanine transaminase (GPT), aspartate transaminase (GOT1 in cytoplasm and GOT2 in mitochondria) and phosphoserine transaminase (PSAT). In these reactions, nitrogen is transferred between glutamate and the α-ketoacids pyruvate (reactions catalyzed by GPT), oxaloacetate (GOT), and phosphohydroxypyruvate (PSAT) to generate alanine, aspartate, and phosphoserine respectively. A fraction of alanine is used for protein synthesis, but many cancer cells in culture also excrete a large amount of alanine (Jain et al., 2012). Aspartate can be generated from the TCA cycle intermediate oxaloacetate in the mitochondria (GOT2) and transported into the cytoplasm to be used in protein and nucleotide synthesis (Birsoy et al., 2015). Notably, a fraction of this aspartate may be converted back into oxaloacetate by the cytoplasmic aspartate transaminase (GOT1), part of the malate-aspartate shuttle. Serine - produced from phosphoserine through PSAT - can be used to produce glycine and precursors for various cellular components, including proteins, nucleotides and lipids, as mentioned previously.

Reactions catalyzed by amino acid transaminases produce α-ketoglutarate. The conversion of glutamate to α-ketoglutarate may also occur through glutamate dehydrogenase (GLUD), releasing the α-nitrogen as free ammonia. The relative contribution of various reactions to α-ketoglutarate
production likely depends on genetic background and micro-environmental factors. For example, KRas-transformed pancreatic cancer cells may rely heavily on aspartate transaminases for metabolism of the glutamine carbon skeleton. When glucose is scarce, glutamate dehydrogenase becomes the predominant route to produce α-ketoglutarate through an AKT/mTORC1 dependent mechanism (Choo et al., 2010). Recently, it has also been demonstrated that GLUD is activated by the mTORC1 pathway through transcriptional repression of SIRT4, an inhibitor of GLUD (Csibi et al., 2013).

Glutamine-derived α-ketoglutarate can serve as an important anaplerotic substrate in the TCA cycle (DeBerardinis et al., 2007). For example, excretion of TCA cycle citrate into cytoplasm to support lipogenesis often depends on α-ketoglutarate-based anaplerosis in rapidly proliferating cells. Metabolism of α-ketoglutarate in the oxidative direction of the TCA cycle generates reducing equivalents for the mitochondrial respiratory chain and contributes to ATP production. In this way, glutamine oxidation serves as a major source of energy in some cancer cells (DeBerardinis et al., 2007).

There is evidence that a portion of the glutamine carbon skeleton can be converted into lactate (DeBerardinis et al., 2007; Ying et al., 2012). One route of glutamine to lactate conversion involves production of glutamine-derived malate, which can be converted to pyruvate via cytosolic malic enzyme (ME1) with the reduction of NADP+ to NADPH. Flux through ME1 can be an important source of NADPH for some cells (DeBerardinis et al., 2007). In K-Ras transformed or SMAD4-deleted pancreatic cancer cells, transport of malate has been proposed to be particularly important for NADPH production (Dey et al., 2017; Son et al., 2013). NADPH is essential for providing
reducing equivalents for nucleotide and lipid synthesis as well as for quenching reactive oxygen species produced during rapid cell proliferation through regeneration of reduced glutathione.

**Aspartate limitation in conditions of electron acceptor deficiency**

Of the other non-essential amino acids, aspartate has received increasing attention in recent years. In addition to its requirement in proteins, aspartate is unique in its contribution to both the carbon backbone of the nucleotide base of pyrimidines, and nitrogens in the nucleotide base of purines. Several recent studies have identified a critical role of aspartate in connection to cells deficient for electron acceptors or in hypoxia (Birsoy et al., 2015; Gui et al., 2016; Sullivan et al., 2015).

The primary nutrients taken up by cancer cells from the environment, glucose and glutamine, are more reduced than many of the precursors for biomass, such as aspartate (Hosios and Matthew, 2018). As a consequence, a series net oxidative reactions are required for aspartate synthesis, in which electrons are donated from reduced substrates and transferred to the cofactor NAD+ to generate its reduced form NADH. Notably, transporters of NAD+/NADH across the cellular membrane do not exist. Therefore, to sustain production of oxidized biomass components, the electrons carried by NADH must be transferred to a terminal electron acceptor. In normoxic conditions, molecular oxygen fulfills this role through complex IV in the mitochondria. Through a series of electron transfer events mediated by the electron transport chain (ETC), oxygen is reduced to water and excreted from the cell at complex IV, while NAD+ is regenerated from NADH at complex I. The regenerated NAD+ can then be further consumed by oxidative reactions, continuing the cycle (Sullivan et al., 2015).
Transport of physiological levels of extracellular aspartate is extremely poor in vitro and in vivo, suggesting that cells must synthesize aspartate de novo (Garcia-Bermudez et al., 2018; Sullivan et al., 2018). Furthermore, human cancer cells do not express the enzyme aspariginase, which deaminites asparagine to form aspartate (Sullivan et al., 2018). Therefore, aspartate synthesis in cancer cells mainly derives from glutamine carbons. As a result, in conditions in which oxygen availability is limiting, such as hypoxia or compromised mitochondrial respiration, cells cannot regenerate sufficient NAD+ through complex I to support synthesis of aspartate. Therefore, aspartate becomes the primary metabolic limitation of these cells. Consistently, it has long been known that cells lacking a functional mitochondria cannot proliferate in vitro (King and Attardi, 1989), and that mitochondrial function is required for tumor progression in vivo (Weinberg et al., 2010). Notably, these cell lacking a functional mitochondria can proliferate in the presence of pyruvate, which serves as an alternative exogenous electron acceptor to enable NAD+ regeneration through LDH and aspartate synthesis (Birsoy et al., 2015; Sullivan et al., 2015). Consistent with aspartate being the primary metabolic limitation of these cells, supplementing supraphysiological levels of aspartate rescues proliferation in vitro and increasing intracellular asparatate levels significantly increases tumor growth in vivo ((Birsoy et al., 2015; Garcia-Bermudez et al., 2018; Sullivan et al., 2015, 2018)).

Lipid acquisition strategies of cancer cells

In addition to amino acids, membrane lipids are a major component of cellular biomass and cancer cells must employ different strategies to increases membrane synthesis rates. Cancer cells are
capable of upregulating de novo lipogenesis, and the sterol regulatory element-binding protein (SREBP) family, the major transcriptional regulators of de novo fatty acid synthesis, are positively regulated by the mammalian target of rapamycin complex 1 (mTORC1), a master coordinator of cellular anabolism and growth (Peterson et al., 2011). Indeed, increases rates of de novo lipogenesis have been observed to regulate the growth of a number of different tumor types (Orita et al., 2008). Cells employ various routes for fatty acid synthesis. In normoxic conditions, glucose is the primary carbon source of cytosolic acetyl-CoA, which undergoes a series of condensation steps by fatty acid synthase (FAS) to generate the saturated fatty acid palmitate. In the case of normoxia, glucose derived citrate is effluxed from the mitochondrial TCA cycle and is subsequently cleaved by ATP-citrate lyase to generate cystosolic acetyl-CoA. However, in conditions of hypoxia or mitochondrial dysfunction, where pyruvate entry into the mitochondria is attenuated, glutamine may serve as an alternative carbon source through the so-called reduction carboxylation route (Metallo et al., 2011; Mullen et al., 2012; Wise et al., 2011). Here, glutamine-derived alpha-ketoglutarate undergoes reductive carboxylation by mitochondrial or cystosolic isocitrate dehydrogenase reactions (IDH1 and IDH2 in the cytosol and mitochondria respectively) to generate citrate in the cytosol. Thus, cancer cells demonstrate plasticity in de novo lipid synthesis strategies following different environmental contexts.

In addition to glucose and glutamine, extracellular acetate may directly increase pools of intracellular acetyl-CoA through the activity of acetyl-CoA synthetase (ACSS2). Interestingly, in low-serum conditions or hypoxia, cancer cells upregulate activity of ACSS2, suggesting that acetate becomes an important carbon source in conditions of metabolic stress (Schug et al., 2015).
Indeed, knockdown of ACSS2 reduced growth in a hypoxia-specific manner, and significantly attenuated growth of tumor xenografts (Schug et al., 2015).

Recently, a number of studies have demonstrated that in addition to increased rates of de novo lipid synthesis, cancer cells are capable of directly acquiring lipids from the extracellular environment. In some cancer cell lines, the measured fraction of lipids acquired exogenously is substantially greater than the amount that is synthesized de novo (Hosios et al., 2016). The exogenous uptake of whole lipids may be further increased in hypoxia, with a preference for monounsaturated fatty acids (Kamphorst et al., 2013). This preferences results from the decreased activity of the stearoyl-CoA desaturate (SCD) enzymes in hypoxia, which directly require oxygen as a substrate, resulting in a decreased membrane desaturation index. To prevent the loss of membrane fluidity, cancer cells can respond by directly taking up exogenous lysophospholipids lipids, thus bypassing the oxygen-dependent SCD reaction (Kamphorst et al., 2013).

Lipids droplets, intracellular organelles containing triglycerides and sterols esters, have also been shown to accumulate in cancer cells under hypoxia (Bensaad et al., 2014). The increase in intracellular lipids occurs via direct uptake from the media through HIF-1α regulation of FABP3/7, proteins important for the long-chain fatty acid transport. Interestingly, hypoxic cells were also shown to decrease de novo lipogenesis, likely through downregulation of SREBP1/2. The increased lipid droplets in hypoxia were shown to contribute to cell growth and survival after reoxygenation (Bensaad et al., 2014).


Wise, D.R., DeBerardinis, R.J., Mancuso, A., Sayed, N., Zhang, X.-Y., Pfeiffer, H.K., Nissim, I.,


CHAPTER 3: QUANTIFYING SPATIOTEMPORAL DYNAMICS AND NOISE IN ABSOLUTE MICROBIOTA ABUNDANCES USING REPLICATE SAMPLING

Author statement

Some passages and figures have been adapted or quoted verbatim from the article: Ji, BW, Sheth, RU, Dixit, Purushottam D, Kaufman, A, Wang, HW, Vitkup, D. Quantifying spatiotemporal dynamics and noise in absolute microbiota abundances using replicate sampling, 2018. In Review.

Authors

Brian W. Ji
1,2†, Ravi U. Sheth
1,2†, Purushottam D. Dixit1, Andrew Kaufman1, Harris H. Wang
1,3* & Dennis Vitkup
1,4*

Author Affiliations

1Department of Systems Biology, Columbia University, New York, NY, USA.
2Integrated Program in Cellular, Molecular, and Biomedical Studies, Columbia University, New York, NY, USA.
3Department of Pathology and Cell Biology, Columbia University, New York, NY, USA
4Department of Biomedical Informatics, Columbia University, New York, NY, USA.
†These authors contributed equally to this work.
Abstract

Metagenomic sequencing has enabled detailed investigation of diverse microbial communities, but understanding their spatiotemporal dynamics remains an important challenge. Here we present DIVERS, a widely applicable method based on replicate sampling and spike-in sequencing that quantifies the contributions of temporal dynamics, spatial sampling variability and technical noise to the variances and covariances of absolute bacterial abundances. We apply DIVERS to high-resolution time series profiling of the human gut microbiome and a spatial survey of a soil bacterial community in Manhattan’s Central Park. Our method reveals complex spatiotemporal dynamics of individual bacteria and unmasks key features of their behavior hidden from previous analyses.

Results

Metagenomic sequencing is widely used to explore patterns of bacterial abundances and the spectrum of functions carried out by diverse microbial communities (Huttenhower et al., 2012; Kozich et al., 2013; Segata et al., 2012; Thompson et al., 2017). However, as research efforts move beyond static descriptions of communities towards understanding their complex spatiotemporal dynamics (Faust et al., 2015; Hunt et al., 2008; Lloyd-Price et al., 2017; Martin-Platero et al., 2018), a number of key challenges remain to be addressed. First, robust and quantitative frameworks are required to characterize both the temporal variability and spatial heterogeneity of bacterial abundances across environments (Tropini et al., 2017; Zhang et al., 2014). Second, the effects of technical noise arising from sample preparation and sequencing must be quantified (Costea et al., 2017; Sinha et al., 2017). Third, measurements of absolute bacterial abundances are necessary to
correct for possible compositional artifacts associated with relative abundances (Friedman and Alm, 2012; Vandeputte et al., 2017). Importantly, quantitative approaches for understanding both the spatiotemporal dynamics and noise profiles of microbial communities must be experimentally tractable, minimizing the number of sample preparations and costs associated with data collection, processing and sequencing.

To address these challenges, we have developed Decomposition of Variance Using Replicate Sampling (DIVERS), a broadly applicable method for metagenomic sequencing studies. DIVERS utilizes the laws of total variance and covariance to provide a principled mathematical approach for separating the contributions of time, spatial sampling location and technical noise to measured abundance variances for individual taxa and covariances for pairs of taxa:

\[
\text{Var}(X_i) = \text{Var}_{\text{Temp}} E(X_i|T) + E_{\text{Sp}} \text{Var}_{\text{Sp}} E(X_i|S,T) + E_{\text{Tech}} \text{Var}(X_i|S,T) \tag{3.1}
\]

\[
\text{Cov}(X_i, X_j) = \text{Cov}_{\text{Temp}} \left( E(X_i|T), E(X_j|T) \right) + E_{\text{Sp}} \text{Cov}_{\text{Sp}} \left( E(X_i|S,T), E(X_j|S,T) \right) + E_{\text{Tech}} \text{Cov}(X_i, X_j|S,T) \tag{3.2}
\]

where \(X_i\) and \(X_j\) denote the absolute abundance of individual bacterial taxa \(i\) and \(j\), \(S\) and \(T\) are space- and time-associated random variables capturing the respective spatial and temporal processes affecting the abundances of taxa \(i\) and \(j\), and \(E\), \(\text{Var}\) and \(\text{Cov}\) denote the expectation, variance and covariance of random variables respectively.
We derived unbiased statistical estimators for each of the six terms in equations (1) and (2), and devised a workflow to enable their calculation directly using experimental measurements (Appendix C, Methods). Importantly, DIVERS requires only two samples obtained from randomly chosen spatial locations at each time point of a longitudinal microbiome study. One of these spatial replicates is then split in half to obtain two technical replicates, and absolute abundance measurements on the resulting three samples are performed using a spike-in procedure during sample processing (Stämmker et al., 2016; Tkacz et al., 2018) (Fig. 3.1a, Appendix C, Methods). The key idea behind this approach is that bacterial taxa with genuine temporal fluctuations should also exhibit large abundance covariances between spatial replicate pairs across time points. Spatial variability, quantified by differences in abundances between the two random locations, and technical noise decrease temporal covariances. Interestingly, our sampling scheme and underlying mathematical model are conceptually similar to the dual reporter approach previously used to separate intrinsic and extrinsic noise in gene expression profiles (Elowitz et al., 2002; Swain et al., 2002) (Appendix C).

We first assessed the performance of DIVERS on synthetic data, where the underlying temporal, spatial sampling and technical contributions to bacterial abundance variances were known. Towards that end, we performed stochastic simulations of bacterial community dynamics that explicitly incorporated spatial abundance heterogeneity, as well as technical noise associated with experimental measurement error (Methods). For all simulated species in the community, DIVERS was able to accurately quantify each of the three variability sources contributing to their observed dynamics (r.m.s. error = 0.02 for fractional variance contributions) (Fig. A1).
To demonstrate the utility of DIVERS in a natural ecological setting, we first explored the human gut microbiome, an ecosystem known to exhibit complex spatiotemporal dynamics (David et al., 2014; Zhang et al., 2014). To that end, we carried out 16S rRNA sequencing of fecal samples collected daily over the course of three weeks from a healthy male individual (Fig. 3.1a, Methods). We verified that our spike-in strain was not endogenous to the gut microbiome (Fig. A2a), and confirmed the accuracy of our spike-in approach to estimate total fecal bacterial loads using serial dilutions of input fecal matter (Fig. A2c, Methods). Furthermore, technical replicate measurements of fecal samples collected over the time series showed excellent reproducibility (Pearson’s r = 0.9) (Fig. A2d). We next characterized total baseline bacterial abundance variation in the human gut microbiome. Consistent with previous results (Vandeputte et al., 2017), we found that total bacterial abundances fluctuated substantially across different samples (coefficient of variation = ~0.5) (Fig. 3.1b). Notably, the observed variability was dominated by daily temporal changes, with total bacterial loads remaining relatively constant across different spatial locations on each day (Fig. 3.1c).

Using measurements of total bacterial loads, we calculated the absolute abundances of all operational taxonomic units (OTUs) and then used DIVERS to decompose the abundance variance of individual OTUs (Methods and Appendix C). Interestingly, variance profiles exhibited two regimes when OTUs were grouped by average abundance, with a transition occurring at ~0.01% in relative abundances (Fig. 3.1d and Fig. A3). Fluctuations of OTUs below this abundance cutoff could be primarily explained by technical sources of variability, generally consistent with Poissonian sampling noise (Fig. A4b,c). In contrast, variability of OTUs above this cutoff largely reflected temporal changes (Fig. 3.1d and Fig. A4a). Differences across spatial sampling locations also contributed a substantial fraction of total variability (on average ~20% for OTUs with an
average absolute abundance > 10$^4$), demonstrating significant spatial heterogeneity of fecal samples.

To further experimentally validate the developed workflow and the variance decomposition model, we performed a set of controlled experiments that specifically eliminated either temporal or spatial variability from fecal samples (Methods). First, we collected fecal samples from ten independent spatial locations of the same stool specimen. This procedure effectively simulated five consecutive time points of the DIVERS protocol, but without any temporal contribution to microbiota variability. Second, to remove spatial variability, we carried out eight consecutive days of sampling with spatial replicates that were homogenized on each day before sequencing (Methods). Reassuringly, the model correctly predicted no temporal or spatial contributions to OTU abundance variability when the corresponding signals were removed from the data (Fig. 3.1e).
Figure 3.1. Variance decomposition of gut bacterial abundances using DIVERS.

(a) Illustration of the DIVERS workflow applied to the fecal microbiome. Samples are collected from two random spatial locations (X and Y from the purple site, Z from the blue site, as shown on the left side of the figure) on each day of microbiome sampling, and two technical replicates (X and Y) are prepared from one of these spatial locations. The resulting three samples from each day (X, Y, and Z) are subjected to a custom spike-in procedure to estimate absolute bacterial abundances. The DIVERS variance decomposition model is then applied (right side of the figure) to abundance profiles of each taxa to quantify contributions of temporal variability, spatial sampling heterogeneity and technical noise to total abundance variability. (b) Temporal profiles of total bacterial densities in the human gut microbiome. X and Y correspond to technical replicate measurements of total bacterial density from a single spatial location, while Z corresponds to a second spatial replicate. Gray line shows the average of spatial replicates. Total bacterial densities are reported in arbitrary units and normalized to a mean of one (Methods). (c) Variance fraction of total bacterial densities attributed to technical (N, purple), spatial sampling (S, blue) and temporal (T, red) factors as calculated by the variance decomposition model. (d) Variance decomposition of individual OTU abundances. Absolute OTU abundances were obtained by multiplying relative abundance profiles by the total bacterial density in each sample and are reported in arbitrary units (Methods). OTUs are binned by their mean abundance across all samples, and stacked bars show the average variance contribution of technical, spatial sampling and temporal sources to OTUs within each bin. Dashed vertical line corresponds to a mean absolute abundance of \(10^{-4}\). Error bars represent the standard error of the mean (SEM). (e) Variance decomposition of microbiota abundances from control experiments. Top: DIVERS applied to stool samples without spatial variability; bottom: DIVERS applied to stool samples without temporal variability.

The ability of DIVERS to unmask temporal and spatial variances of individual OTUs makes it possible to investigate microbiota fluctuations in the human gut (Ma, 2015) using absolute bacterial abundances. Interestingly, temporal variances calculated from DIVERS showed a robust power law dependence on average OTU abundance, following a relationship known as Taylor’s law in ecology (Taylor, 1961) (Fig 3.2a, power law exponent \(b = 1.9\), Pearson’s \(r = 0.97\) on log-transformed means and variances). In addition, DIVERS allowed us to investigate the relationship between average absolute abundances and their spatial variances, which could also be well-
described by a power law (Fig 3.2b, $b = 1.7$, $r = 0.96$). These results suggest that, in contrast to the null model of randomly distributed OTU abundances across time and space ($b = 1$), bacterial species in the gut microbiome display significantly more complex spatiotemporal dynamics (Kilpatrick and Ives, 2003; Zhang et al., 2014).

Beyond general ecological relationships, we used our approach to identify specific taxa with particularly high spatial or temporal contributions to their total abundance variance (Fig 3.2c, Table A3, Methods). Interestingly, the time series of several high-abundance OTUs showed behavior primarily shaped by either spatial (OTU 13, Genus: Clostridium IV and OTU 122, Genus: Terrisporobacter) or temporal variation (OTU 12, Genus: Bifidobacterium and OTU 25, Genus: Lachnospiracea incertae sedis) (Fig 3.2d). These examples demonstrate that DIVERS may be used to characterize the spatiotemporal dynamics of individual OTUs in the human gut.
Fluctuations in bacterial abundances often result from the collective behavior of multiple different taxa, whose interactions are reflected in correlated abundance changes (Faust and Raes, 2012). DIVERS can also be used to quantify the factors contributing to abundance correlations between pairs of OTUs in a microbial community (Methods and Appendix C). Applying this analysis to the human gut microbiome, we found that the majority of pairwise abundance correlations were due to temporal sources, with relatively smaller contributions from spatial sampling location and technical noise (Fig. 3.3a and Fig. A5a,b). Consistent with previous results (Vandeputte et al., 2017), we also found that total correlations based on absolute abundance measurements were generally larger than correlations calculated using relative abundances, an effect primarily caused by the variance in total bacterial loads across samples (Fig. A5c,d and Appendix C).

Next, we examined factors contributing to the correlations of OTU abundances within and between the four most prevalent bacterial phyla in the human gut. Interestingly, the Bacteroidetes exhibited significantly larger intra-phyla temporal abundance correlations compared to the rest of the
community (p < 1e-10, Wilcoxon rank sum test) (Fig. 3.3b, Fig. A6). This result was also observed at the family level, and was not due to differences in 16S rRNA sequence similarity across taxa (Fig. A7, Fig. A8). The coordinated temporal changes of *Bacteroidetes* in the human gut may reflect fluctuations in the availability of dietary polysaccharides on each day that are specifically metabolized by these bacteria (Sonnenburg et al., 2010, 2016), as well as previously observed cross-feeding interactions between these taxa (Rakoff-Nahoum et al., 2014, 2016). In addition, our analysis revealed several interesting examples of OTU pairs with positive and negative correlation contributions from temporal and spatial factors (Fig. 3.3c,d). These examples highlight the diversity of bacterial dynamics in the gut, and demonstrate the ability of DIVERS to disentangle the factors contributing to abundance correlations between different taxa.

**Figure 3. 3.** Decomposition of pairwise OTU abundance correlations in the human gut microbiome.
(a) Boxplots of total, temporal, spatial and technical correlations for all pairs of abundant OTUs (average absolute abundance > 10^{-4}). Boxes denote the median and interquartile ranges, with maximum whisker lengths three times the interquartile range. (b) Temporal correlations of OTU abundances within and between different phyla. Colors reflect average temporal correlations between pairs of OTUs belonging to the indicated phyla. Data are shown for all highly abundant OTUs (mean absolute abundance >10^{-4}) from the Actinobacteria (n = 10), Bacteroidetes (n=15), Firmicutes (n=103), and Proteobacteria (n=5). (c) Temporal and spatial correlations for all pairs of abundant OTUs (average absolute abundance > 10^{-4}). Colored points (1-3) indicate pairs of OTUs with temporal profiles shown in d. (d) Temporal abundance profiles for pairs of OTUs highlighted in c. Pairs exhibit (from left to right): 1) Substantial negative temporal ($\rho_T = -0.63, p = 4e-4$), 2) substantial positive spatial ($\rho_S = 0.85, p = 3e-4$), and 3) substantial positive temporal ($\rho_T = 0.90, p < 1e-4$) correlations. For every OTU pair, blue and pink solid lines show abundances of each OTU measured from the same spatial location (Z). Blue and pink dashed lines show the average between technical replicates ($1/2(X+Y)$) of each OTU measured from the second spatial location. Note that the large spatial correlation between OTUs 13 and 33 (panel 2) is reflected in similar profiles of the two dashed lines, as well as the two solid lines (measurements from the same spatial location); the lower temporal correlation between these OTUs is reflected in more dissimilar profiles of solid and dashed lines of the opposite color (measurements from different spatial locations).

The DIVERS variance decomposition framework is not limited to the human gut microbiome, and can be applied to understand patterns of bacterial abundance fluctuations across space and time in diverse environmental settings. To demonstrate this, we conducted a spatial survey of a soil microbial community in Manhattan’s Central Park in New York City. To explore spatial abundance variability, we utilized a modified protocol that inverted the hierarchy of spatial and temporal sampling replicates (Appendix C). Specifically, we collected soil samples from twenty-eight sites located roughly uniformly around a small man-made pond located in the northwest section of Central Park (Fig. 3.4a, Fig. A9, Table A4, Methods). Samples were collected from identical locations at two time points one week apart in June of 2018. Following the DIVERS protocol, one of the temporal replicates from each spatial location was subjected to two independent rounds of sample preparation and sequencing (Fig. 3.4a), and a similar spike-in approach was used to estimate total bacterial densities in each soil sample (Fig. A2c, Methods).
The soil community analysis revealed substantial variability in total bacterial densities across soil samples (CV ~ 0.4), which could be primarily attributed to spatial variability (Fig. 3.4b, Fig. A10a). We next applied the DIVERS variance decomposition model to absolute soil OTU abundances. Similar to the human gut, we found that technical noise predominated in OTUs with lower abundances ($\log_{10}$ mean absolute abundance $<-4.5$). For OTUs with higher abundances ($\log_{10}$ mean absolute abundance $>-4.5$), spatial sampling location was the major source of variability (Fig. 3.4c, Fig. A10). Temporal variability, defined over a one week time scale, contributed a relatively smaller fraction of the total variance. Notably, when applying the DIVERS covariance decomposition model to soil bacteria, we observed a relatively low degree of abundance correlations across all OTUs, as well as within and between different taxa (Fig. A11). These results indicate relatively weak patterns of spatial co-occurrences of soil bacteria (O’Brien et al., 2016), but significant spatial abundance heterogeneity, likely due to differing, stable environmental niches across soil sites (Fierer and Jackson, 2006; O’Brien et al., 2016). The soil community analysis further demonstrates the ability of DIVERS to identify different patterns of bacterial abundance variation and covariation across diverse ecosystems.
Figure 3. Variance decomposition applied to the dynamics of soil bacterial abundances.

(a) Illustration of the DIVERS sampling protocol applied to a Central Park soil microbial community. Samples were collected one week apart from twenty-eight sites located around a small pond in the northwest area of the park. At each site \(i\), two technical replicates were prepared using samples from one time point (\(X_i\) and \(Y_i\)), whereas a single measurement was made at the other time point (\(Z_i\)).

(b) Absolute bacterial densities of soil samples across different spatial locations. \(X\) and \(Y\) correspond to measurements made at a single time point, while \(Z\) corresponds to measurements from the second time point.

(c) Variance decomposition of individual OTU abundances. Absolute OTU abundances were obtained by multiplying relative abundance profiles by the total bacterial density in each sample and are reported in arbitrary units (Methods). OTUs are binned by their mean abundance across all samples, and stacked bars show the average variance contribution of technical, spatial sampling and temporal sources to OTUs within each bin. Error bars represent the SEM.

Conclusion

While current sequencing technologies enable bacterial communities to be profiled at high temporal resolution, novel approaches are required to reveal key features of ecosystem dynamics.

Our results demonstrate the ability of DIVERS to quantify both the spatiotemporal dynamics and
noise profiles of diverse microbial communities, while requiring only a small number of additional samples compared to current metagenomic sequencing protocols. Although we focus on human gut and soil microbial communities in this study, DIVERS can be readily applied to explore patterns of variation in any bacterial ecosystem across different hosts and environments (Appendix C). Moreover, given the flexibility of the developed quantitative framework, it can be easily extended to other sequencing-based applications, such as the characterization of human immune cell repertoires (Wargo et al., 2016) and gene expression changes in tumors (Tirosh et al., 2016).
Methods

**Ethical review.** This study was approved and conducted under Columbia University Medical Center Institutional Review Board protocol AAAR0753. Written informed consent was obtained from the subject in the study, a healthy male adult.

**Fecal sample collection and storage.** Fecal samples were collected daily over the course of twenty days, with two additional samples taken on days 27 and 48 of the study. After defecation, inverted sterile 200 µL pipette tips (Rainin RT-L200F) were used to core out a small sample from the stool specimen, which was placed immediately in a sterile cryovial (Sarstedt 72.694.106). Samples were then immediately placed in a -20 °C freezer and transferred to a -80 °C freezer for long-term storage.

**Replicate fecal sampling experimental protocol.** To enable decomposition of gut bacterial abundance variability into temporal, spatial and technical contributions, a replicate sampling approach was utilized. Specifically, on each day of the time series, two fecal samples were collected from random spatial locations of the same stool specimen. For one of these samples, two technical replicates were prepared in parallel by splitting the individual fecal core. Thus, a total of three samples were processed for each day of the time series: two technical replicates from a single spatial location (denoted samples X and Y) and a second spatial replicate (denoted sample Z). To further characterize technical noise, a single fecal sample was also subjected to 12 independent rounds of sample processing and sequencing. Metadata associated with all fecal samples are
provided in Table A1. Theoretical details associated with the DIVERS approach are described in the Appendix C.

**Soil site description, sample collection and storage.** Soil samples were collected in June of 2018 from The Pool in Central Park, Manhattan (approximately 40.795°N, 73.960°W), a man-made body of water located in the northwest area of the park. Soil cores were collected on two days exactly one week apart from twenty-eight sites located on the periphery of the water’s edge. The average distance between adjacent sites was ~8 meters. Photographs were taken at each site to ensure sampling accuracy at the same location from different time points (Fig. A9). Following soil collection, samples were transferred to a -80 °C freezer for long-term storage.

**Replicate soil sampling experimental protocol.** Similar to our fecal sampling protocol, a replicate sampling approach was utilized to collect soil bacteria. At each of the twenty-eight soil sampling sites, technical replicates were prepared from a single sample collected at one of the two time points by splitting the individual soil core. The time points for which technical replicates were prepared were alternated between neighboring sites. Therefore, three samples were processed at each of twenty-eight sites in the environment: two technical replicates from a single time point (denoted samples X and Y) and a second temporal replicate (denoted sample Z). Metadata associated with all soil samples are provided in Table A4. Due to technical error associated with sample preparation, sites 2 and 6 were excluded from any further downstream analyses. Theoretical details associated with the DIVERS approach are described in the Appendix C.
**Spike-in strain for calculation of bacterial absolute abundances.** A spike-in approach was utilized during sample processing to allow for calculation of total bacterial abundances per mass of fecal or soil matter. Sporocarcina pasteurii (ATCC 11859), an environmental bacterium that was confirmed to be absent in our fecal and soil samples, was grown to saturation in NH4-YE medium (ATCC medium 1376). It was then concentrated by centrifugation, resuspended in ~0.1X volume phosphate buffered saline (PBS) with 20% glycerol, and stored in cryovials at -80 °C for subsequent use during genomic DNA extraction.

**Sample genomic DNA extraction.** Genomic DNA (gDNA) extraction was performed using a custom liquid handling protocol based on the Qiagen MagAttract PowerMicrobiome DNA/RNA Kit (Qiagen 27500-4-EP) adapted for lower volumes. Briefly, a 96 well plate (Axygen P-DW-20-C) was loaded with 1 mL of 0.1 mm Zirconia Silica beads (Biospec 11079101Z) using a loading device (Biospec 702L). During sample processing, appropriate negative controls were run on each plate (i.e. water control). 10 uL of thawed and concentrated spike-in strain was added to each well; for soil samples, the spike-in strain was diluted 1:25. 10-500 mg of each sample (average 45.9 mg, standard deviation 14.7 mg for fecal samples; average 298.5 mg, standard deviation 62.8 mg for soil samples) was added to the plate using a sterile plastic spatula, and the weight added for each sample was determined via an analytical balance. 750 µL of lysis solution was then added to each well (90 mL master mix, 9 mL 1M Tris HCl pH 7.5, 9 mL 0.5M EDTA pH 8.0, 11.25 mL 10% SDS, 22.5 mL Qiagen lysis reagent, 38.25 mL nuclease free water). The plate was centrifuged down for 1 min at 4500xg and a bead sealing mat was affixed to the plate (Axygen AM-2ML-RD). The plate was then placed on a bead beater (Biospec 1001) and subjected to bead beating for 5 min followed by 10 min for cooling. This bead beating cycle was repeated, for a total of 10 min of bead
beating. The plate was centrifuged down for 5 min at 4500xg and 200µL of supernatant was transferred to a V-bottom microplate. 35 µL of Qiagen inhibitor removal solution was added to each well and mixed by vortexing, incubated 4 °C for 5 min, and the plate was again centrifuged down for 5 min at 4500xg. 100 µL of supernatant was removed from the plate and placed in a round-bottom plate (Corning 3795). The plate was then placed on a robotic liquid handler (Biomek 4000) for magnetic bead purification of the supernatant per the manufacturers recommendations but at a scaled volume; magnetic beads in binding solution were mixed in each well, and subjected to 3 washes with wash solution and elution in 100 µL of nuclease free water into a new plate.

**16S rRNA amplicon sequencing.** 16S sequencing of the V4 region was performed utilizing a custom protocol and a dual indexing scheme adapted from Kozich et al(Kozich et al., 2013). Briefly, dual indexing sequencing primers were adapted from the previous study, but we utilized Illumina Nextera barcode sequences and altered 16S primers to match updated 505f and 806rB primers (see Table S2 for sequences). A 20 µL PCR amplification was set up in a 96 well skirted PCR microplate: 1 µM forward 5XX barcoded primer, 1 µM reverse 7XX barcoded primer, 1 µL prepared gDNA, 10 uL NEBNext Q5 Hot Start HiFi Master Mix (NEB M0543L), 0.2X final concentration SYBR Green I. A quantiative PCR amplification (98°C 30s; cycle: 98°C 20s, 55°C 20s, 65°C 60s, 65°C 5m) was performed and cycling was stopped during exponential amplification (typically 12-20 cycles) and the reaction was advanced to the final extension step.

The resulting PCRs were quantified utilizing a SYBR Green I dsDNA assay; 2 µL of PCR product was added to 198 µL of TE with 1X final concentration SYBR Green I and fluorescence was quantified on a microplate reader. Samples were pooled based on this quantification on a robotic
liquid handler (Biomek 4000). The resulting ~390 bp amplicon from the pool was then gel-purified utilizing a 2% E-gel (Invitrogen) and Wizard SV gel extraction kit.

Final libraries were then quantified by Qubit dsDNA HS assay and sequenced on the Illumina MiSeq platform (V2 500 or 300 cycle kit) according to the manufacturers instructions with modifications. Specifically, the library was loaded at 10 pM with 20% PhiX spike-in, and custom sequencing primers were spiked into the MiSeq reagent cartridge (6 uL of 100 µM stock; well 12: read1, well 13: index1, well 14: read2).

**Sequence analysis and OTU clustering.** Resulting sequence data was analyzed with the USEARCH (Edgar, 2010) pipeline. Specifically, raw reads were merged using the –fastq_mergepairs command (for 2x250 reads, the options –fastq_maxdifs 10 –fastq_maxdiffpct 10 were utilized). Merged sequences were filtered using the –fastq_filter command with options –fastq_maxee 1.0 and –fastq_minlen 240. Resulting sequences were dereplicated (–drep_fulllength), clustered into OTUs (–cluster_otus) and the merged reads were searched against OTUs sequences (–usearch_global) at 97% identity. Taxonomic assignments of OTUs were made using the RDP classifier (Wang et al., 2007).

**Calculation of OTU absolute abundances.** Total bacterial densities in each sample were calculated using the following formula:

\[
R_i = \frac{C_0}{C_0 + \rho_i W_i}
\]
where, $R_i$ is the sequenced relative abundance of the spike-in strain in sample $i$, $C_0$ is the constant amount of spike-in strain (units of total DNA copies) added to each sample, $W_i$ is the weight of the fecal or soil sample $i$ (mg), and $\rho_i$ is the total bacterial density per fecal/soil mass (DNA copies/mg). Solving for $\rho_i$,

$$\rho_i = \frac{C_0(1 - R_i)}{R_i W_i}$$

where we have measured $R_i$ and $W_i$ experimentally. Note that relative changes in $\rho_i$ are independent of the constant $C_0$. We therefore scaled total bacterial densities within fecal or soil samples to a mean of unity. Relative abundance profiles (with the spike-in strain excluded) were then multiplied by this scaled quantity to obtain absolute OTU abundances in arbitrary units that were used for all analyses.

**Assessment of the DIVERS spike-in sequencing approach to estimate absolute bacterial abundances.** To assess the accuracy of the DIVERS spike-in approach in estimating absolute abundances, we performed a spike-in dilution series. Specifically, two fecal samples from different individuals were homogenized in 5X volume sterile PBS by vortexing, and passed through a 40 micron sterile filter. The fecal filtrate was then serially diluted 1:2 in sterile PBS to generate samples with exponentially decreasing fecal matter. Constant volumes (100 uL) of the undiluted and diluted samples were then subjected to the DIVERS spike-in sequencing approach as described previously.
Based on the above formula used to calculate total bacterial densities, we derived a single relationship that described the expected behavior of sequenced spike-in strain abundances across the dilution series:

\[
\frac{R_0(1 - R_i)}{R_i(1 - R_0)} = 2^{-i}
\]

where \(R_0\) is the sequenced relative abundance of the spike-in strain in the original, undiluted fecal sample, and \(R_i\) is the relative abundance of the spike-in strain in the \(i\)th sample of the dilution series (i.e. sample \(i = 1\) contains one half of the input fecal matter of the original sample). We show excellent agreement between expected and observed behavior in Figure C.2.

**Variance decomposition of OTU abundances and total bacterial loads.** DIVERS utilizes the replicate sampling and sequencing protocol described above to decompose measured bacterial abundance variances. Let \(X\) denote the total bacterial density in a collected sample or the abundance of an individual OTU. Using the law of total variance, the variance of \(X\) can be written as a sum of three components associated with temporal, spatial and technical factors contributing to changes in \(X\) across samples:

\[
Var(X) = Var_T E_S Var(X|S,T) + E_T Var_S Var(X|S,T) + E_T E_S Var(X|S,T)
\]

(3.1A)
where, $S$ and $T$ are space and time-associated random variables capturing the spatial and temporal processes that influence the abundance of $X$ across samples. Following the notation in Fig. 3.1, each of the terms in (1) is estimated as follows (see Appendix C for full derivations):

$$Var_T E_{S|T} E(X|S,T) = Cov(X,Z) \quad (3.2A)$$

$$E_T Var_{S|T} E(X|S,T) = Cov(X - Z, Y) \quad (3.3A)$$

$$E_T E_{S|T} Var(X|S,T) = \frac{1}{2} Var(X - Y) \quad (3.4A)$$

where $X, Z$ and $Y, Z$ denote pairs of spatial replicate measurements of either total bacterial density or individual OTU abundances. As described above, spatial replicates are obtained from two independent spatial locations in the environment at every time point. In contrast, $X$ and $Y$ denote technical replicates that are measured from the same spatial location.

**Covariance decomposition of OTU abundances.** Using the law of total covariance, the covariance between the abundances of any two OTUs $i$ and $j$, denoted $X_i$ and $X_j$, can also be written as a sum of temporal, spatial and technical contributions:

$$Cov(X_i, X_j) = Cov_T \left( E(X_i|T), E(X_j|T) \right) + E_T Cov_{S|T} \left( E(X_i|S,T), E(X_j|S,T) \right) + E_T E_{S|T} Cov(X_i, X_j|S, T) \quad (3.5A)$$
Each of the terms in (5) is estimated using the replicate sampling and sequencing protocol as follows (see Appendix C for full derivations):

\[
\text{Cov}_T \left( E(X_i|T), E(X_j|T) \right) = \text{Cov}(X_i, Z_j)
\]

(3.6A)

\[
E_T \text{Cov}_{S|T} \left( E(X_i|S,T), E(X_j|S,T) \right) = \text{Cov}(X_i - Z_i, Y_j)
\]

(3.7A)

\[
E_T E_{S|T} \text{Cov}(X_i, X_j|S,T) = \frac{1}{Z} \text{Cov}(X_i - Y_i, X_j - Y_j)
\]

(3.8A)

where, \(X_i, Z_i\) and \(Y_i, Z_i\) denote spatial replicate measurements of the abundance of OTU \(i\), and \(X_i, Y_i\) denote technical replicates. To obtain temporal, spatial and technical correlations shown in Fig. 3.2, we normalize each covariance contribution by the respective standard deviations of individual OTUs:

\[
Cor(X_i, X_j) = \frac{\text{Cov}_T(E(X_i|T), X_j|T)}{\sigma_{X_i} \sigma_{X_j}} + \frac{E_T \text{Cov}_{S|T}(E(X_i|S,T), E(X_j|S,T))}{\sigma_{X_i} \sigma_{X_j}} + \frac{E_T E_{S|T} \text{Cov}(X_i, X_j|S,T)}{\sigma_{X_i} \sigma_{X_j}}
\]

(3.9A)

Variances and covariances of OTU abundances were calculated using data obtained across the twenty consecutive days of fecal sampling and twenty-six soil sites. The variance decomposition of total fecal bacterial densities also included samples taken from days 27 and 48 of the times.
series. To minimize artifacts due to technical noise, only OTUs with a log ten mean absolute
abundance >-4 and >-3.5 were included in the covariance decomposition analysis of fecal and soil
samples respectively. These cutoffs was chosen based on the observed variance profiles of
individual OTUs. To compare contributions across gut bacterial phyla, 16S rRNA sequence-based
phylogenetic distances were calculated using the pairwise2 module of Biopython.

**Stochastic simulations of microbiota dynamics.** To assess the performance of the DIVERS
variance decomposition model, we carried out stochastic simulations of bacterial dynamics and
measurement noise. We considered a community of interacting species on a 2D lattice, where at
each time point, species were allowed increase their abundance through birth, decrease their
abundance through death, or migrate randomly to a neighboring location. These dynamics were
governed by the following set of reactions:

\[
N_{x,y}^{(l)} \xrightarrow{b(l)_{x,y}} N_{x,y}^{(l)} + 1 \quad (3.10A)
\]

\[
N_{x,y}^{(l)} \xrightarrow{d(l)_{x,y}} N_{x,y}^{(l)} - 1 \quad (3.11A)
\]

\[
N_{x,y}^{(l)} \xrightarrow{v(l)_{x,y}} N_{x,y}^{(l)} - 1 \quad (3.12A)
\]

\[
N_{x\pm1,y\pm1}^{(l)} \xrightarrow{v(l)_{x,y}} N_{x\pm1,y\pm1}^{(l)} + 1 \quad (3.13A)
\]
where $N_{x,y}^{(i)}$ represents the abundance of species $i$ at grid location $(x, y)$, and $b_{x,y}^{(i)}$, $d_{x,y}^{(i)}$, and $v_{x,y}^{(i)}$ are the respective per-capita birth rates, death rates and migration rates of species $i$ at location $(x, y)$. Migration rates for each species were chosen to be independent of spatial location. Per-capita birth and death rates were given by the following density-dependent logistic equation (Kilpatrick and Ives, 2003):

$$
\mu_{x,y}^{(i)} = r_i \left( 1 - \frac{N_{x,y}^{(i)}}{K_{x,y}^{(i)}} + \sum_{j \neq i} A_{ij} N_{x,y}^{(j)} \right)
$$

(3.14A)

$$
b_{x,y}^{(i)} = \begin{cases} 
\mu_{x,y}^{(i)} & \text{if } \mu_{x,y}^{(i)} > 0 \\
0 & \text{if } \mu_{x,y}^{(i)} < 0 
\end{cases}
$$

(3.15A)

$$
d_{x,y}^{(i)} = \begin{cases} 
\mu_{x,y}^{(i)} & \text{if } \mu_{x,y}^{(i)} < 0 \\
0 & \text{if } \mu_{x,y}^{(i)} > 0 
\end{cases}
$$

(3.16A)

where $r_i$ is the intrinsic growth rate of species $i$, $K_{x,y}^{(i)}$ is the carrying capacity of species $i$ at location $(x, y)$, and $A$ is a matrix encoding competitive interactions between community members (elements of $A$ are positive). To incorporate environmental stochasticity into our model, we multiplied species abundances by a Gaussian random variable at each time step: $N(t + \Delta t)_{x,y}^{(i)} = N(t)_{x,y}^{(i)} \zeta^{(i)}$, with $\zeta^{(i)} \sim N(1, \epsilon)$. Finally, to simulate technical noise associated with experimental measurement error (Grün et al., 2014; Marioni et al., 2008), we modeled final observed abundances as a Poisson random variable $X_{x,y}^{(i)} \sim Poiss\left(N_{x,y}^{(i)}\right)$ with mean and variance equal to $N_{x,y}^{(i)}$. 

65
Simulations were carried out using the Gillespie algorithm on a 10 x 10 lattice with continuous boundary conditions. The following parameters were used for simulations: \( n_{\text{species}} = 10, \nu_i \sim \text{unif}(0.5,2), r_i \sim \text{unif}(0.2,0.5), A_{ij} \sim \text{unif}(0.1,0.5), \epsilon = 3 \times 10^{-5} \). Carrying capacities across grid locations were modeled for each species using a gamma distribution with mean and variance chosen randomly from the distributions \( K_{\text{means},i} \sim \text{unif}(1000,5000) \) and \( K_{\text{variances},i} \sim \text{unif}(5000,20000) \) The true temporal abundance variance \( \sigma_T^2(i) \) for each species was calculated empirically as: 
\[
\sigma_T^2(i) = \frac{1}{T-1} \sum_{t=1}^{T} \left( \langle N(t)^{(i)} \rangle - \langle N(i)^{(i)} \rangle \right)^2,
\]
where \( \langle N(t)^{(i)} \rangle \) is the average abundance of species \( i \) at time \( t \) across spatial locations, \( \langle N(i)^{(i)} \rangle \) is the average abundance of species \( i \) over all time points and all spatial locations, and \( T \) is the length of the simulation. Similarly, the spatial abundance variance of each species was calculated empirically as: 
\[
\langle \sigma_S^2 \rangle_T = \frac{1}{T-1} \sum_{S=1}^{S} \sum_{x,y} \left[ N(t)^{(i)}_{x,y} - \langle N(t)^{(i)} \rangle \right]^2,
\]
where \( S \) is the number of considered spatial locations in the environment. Finally, technical variance was calculated empirically as 
\[
\langle \sigma_N^2 \rangle_{S,T} = \frac{1}{T-1} \sum_{S=1}^{S} \sum_{x,y} \frac{1}{N-1} \sum_{n=1}^{N} \left[ X(t)^{(i)}_{x,y} - N(t)^{(i)}_{x,y} \right]^2 = \langle R(i)^{(i)} \rangle\ (\text{Appendix C}).
\]

Taking \( X(t)^{(i)}_{x,y} \) and \( Y(t)^{(i)}_{x,y} \) to be technical replicates draw from the same spatial location at each time point \( t \), and \( Z(t)^{(i)}_{x,y} \) to be a single technical replicate drawn from a different spatial location, we then used equations (2-4) of the DIVERS variance decomposition model to estimate the temporal, spatial sampling and technical abundance variances of each species in the simulated community. In Fig. A1, we compare these estimated variances using DIVERS to the quantities calculated empirically as described above.
Identification of OTUs with high temporal or spatial variance contributions in the human gut. To minimize effects of technical noise, OTUs were first filtered by abundance (mean absolute abundance >10^{-4}). Of the remaining OTUs, those with temporal variance above 80% or spatial variance above 60% of total variability were identified and given in Table A3.

Removal of temporal or spatial variability from fecal samples. We conducted two sets of control experiments to remove either temporal or spatial variability of OTU abundances from fecal samples. Specifically, to eliminate temporal contributions, we re-sampled a single stool specimen ten times total to simulate five consecutive days of time series sampling. To eliminate spatial variability, replicate sampling was conducted for eight consecutive days; on each day, fecal samples obtained from random spatial locations were homogenized together by combining fecal samples, and then mechanically homogenizing in 1X PBS with a P200 pipette tip. The resulting homogenized sample was then split into technical triplicates and processed following the normal DIVERS protocol.
References


Swain, P.S., Elowitz, M.B., and Siggia, E.D. (2002). Intrinsic and extrinsic contributions to


CHAPTER 4: MACROECOLOGICAL RELATIONSHIPS IN THE DYNAMICS OF GUT MICROBIOTA

Author statement

Some passages and figures have been adapted or quoted verbatim from the article:

Authors

Brian W. Ji\textsuperscript{1,3}, Ravi U. Sheth\textsuperscript{1,3}, Purushottam D. Dixit\textsuperscript{1} & Dennis Vitkup\textsuperscript{1,2*}

Author affiliations

\textsuperscript{1}Department of Systems Biology, Columbia University, New York, NY, USA. \textsuperscript{2}Department of Biomedical Informatics, Columbia University, New York, NY, USA. \textsuperscript{3}These authors contributed equally.
Abstract

The gut microbiome is now widely recognized as a dynamic ecosystem that plays an important role in health and disease (Costello et al., 2012). While current sequencing technologies make it possible to estimate relative abundances of host-associated bacteria over time (Caporaso et al., 2011a; Gohl et al., 2016), the biological processes governing their dynamics remain poorly understood. Therefore, as in other ecological systems (Azaele et al., 2016; Marquet et al., 2005), it is important to identify quantitative relationships describing global aspects of gut microbiota dynamics. Here we use multiple high-resolution time series data obtained from humans and mice (Caporaso et al., 2011b; Carmody et al., 2015; David et al., 2014a) to demonstrate that despite their inherent complexity, gut microbiota dynamics can be characterized by several robust scaling relationships. Interestingly, these patterns are highly similar to those previously observed across diverse ecological communities and economic systems, including the temporal fluctuations of animal and plant populations (Azaele et al., 2006; Keitt and Stanley, 1998; Keitt et al., 2002; Sun et al., 2015) and the performance of publicly traded companies (Stanley et al., 1996). Specifically, we find power law relationships describing short- and long-term changes in gut microbiota abundances, species residence and return times, and the connection between the mean and variance of species abundances. The observed scaling relationships are altered in mice receiving different diets and affected by context-specific perturbations in humans. We use these macroecological relationships to reveal specific bacterial taxa whose dynamics are significantly affected by dietary and environmental changes. Overall, our results suggest that a quantitative macroecological framework will be important for characterizing and understanding complex dynamical processes in the gut microbiome.
Results

The dynamics of gut bacteria can now be monitored with high temporal resolution using 16S rRNA amplicon sequencing (Faust et al., 2015). Recent longitudinal studies have revealed significant day to day variability and marked long-term stability of gut microbiota (Caporaso et al., 2011b; David et al., 2014a; Dethlefsen and Relman, 2011; Faith et al., 2013). Several studies have also identified important factors, such as host diet and lifestyle, that contribute to temporal changes in species abundances (Carmody et al., 2015; David et al., 2014b, 2014a; Smits et al., 2017). However, in contrast to other macroscopic ecological communities, statistical relationships describing gut microbiota dynamics are not well understood. While ideas from theoretical ecology have been applied to understand static patterns of gut microbial diversity and species abundance distributions (Li and Ma, 2016; Shoemaker et al., 2017), a comprehensive macroecolgical analysis of gut microbiota dynamics is currently missing. Therefore, we sought to investigate dynamical relationships in the gut microbiome using several of the longest and most densely-sampled longitudinal studies in humans and mice (Caporaso et al., 2011b; Carmody et al., 2015; David et al., 2014a). The considered data spanned three independent investigations, utilizing different sample collection procedures and sequencing protocols; bacterial abundances in these studies were tracked daily for several weeks in mice and up to a year in humans. Our analysis included four healthy human individuals (A, B, M3, F4) and six individually-housed mice fed either a low-fat, plant polysaccharide (LFPP) diet or a high-fat, high-sugar (HFHS) diet. We use these data to explore the short-term abundance changes and long-term drift of gut microbiota, species residence and return times, and the temporal variability of individual bacterial taxa across humans and
different mouse diet groups. Collectively, our study provides a comprehensive characterization of macroecological dynamics in the gut microbiome.

Following a quantitative framework used previously to examine the ecological dynamics of animal populations (Keitt and Stanley, 1998; Keitt et al., 2002), we first investigated short-term temporal fluctuations of gut microbiota abundances. One of the most basic descriptors of bacterial population dynamics is the daily growth rate, defined as the logarithm of the ratio of consecutive daily abundances, \( \mu_k(t) = \log \left( \frac{X_k(t + 1)}{X_k(t)} \right) \), where \( X_k \) is the relative abundance of a bacterial operational taxonomic unit (OTU) \( k \) at time \( t \). In contrast to the instantaneous growth rate, \( \mu_k(t) \) defines the rate of change of OTU abundances averaged over the course of an entire day. Interestingly, we found that the probability of \( \mu \) averaged over all OTUs closely followed a Laplace distribution, with a characteristic tent shape in log-transformed probabilities (Fig. 4.1a-c). Laplace distributions were highly similar within and between individual humans, and between humans and mice (parameter \( b = 0.73 \pm 0.07, b = 0.82 \pm 0.1 \); mean ± s.d. across all humans and LFPP mice respectively), indicating the universality of these relationships. Moreover, the Laplace distribution described well the daily growth rates of every gut microbiome time series we analyzed, including those defined at various taxonomic resolutions (Fig. B1). In contrast to a Gaussian growth rate distribution, which is expected for bacterial growth affected by random multiplicative processes (Mitzenmacher, 2003; Shoemaker et al., 2017), the Laplace distribution indicates substantially higher probabilities for large short-term bacterial abundance fluctuations. Interestingly, very similar growth rate distributions have been observed across many diverse ecological and economic systems including bird communities (Keitt and Stanley, 1998; Keitt et al., 2002), fish populations (Sun et al., 2015), tropical rain forests (Azaele et al., 2006), publicly
traded company sales (Stanley et al., 1996), and country GDPs (Plerou et al., 1999) (Fig. B2a). Similar to these complex ecological and interacting systems, the gut microbiome may exhibit sudden large-scale abundance fluctuations.

In complex ecosystems, species growth rate distributions often depend on their current abundance (Keitt et al., 2002; Plerou et al., 1999; Stanley et al., 1996). We therefore investigated the relationship between the standard deviation of daily bacterial growth rates and abundances. This analysis revealed that the daily growth rate variability of gut bacteria decreased approximately linearly with increasing mean daily abundances (Fig. 4.1d-f). Moreover, the observed behavior was similar between human and mouse gut microbiomes (regression slopes $r = -0.15 \pm 0.01$, $-0.17 \pm 0.03$; mean ± s.d. across humans and mice). Thus, likely due to the presence of more stable nutrient niches, highly abundant bacteria exhibit substantially smaller relative day to day fluctuations compared to bacteria with lower abundances.
Figure 4.1. Daily changes in the abundances of gut microbiota.

(a-c) Daily growth rates were defined as $\mu_k(t) = \log \left( \frac{X_k(t+1)}{X_k(t)} \right)$, where $X_k$ is the relative abundance of a given OTU $k$ on day $t$. The distribution of $\mu$ averaged over all OTUs displays a Laplace form, $p(\mu) = \frac{1}{2b} \exp \left( - \frac{|\mu|}{b} \right)$, appearing as a characteristic tent shape in log-transformed probabilities. Results are shown for two individuals from different human studies (A and M3) and mice fed a low-fat plant polysaccharide-based (LFPP) diet. Laplace exponents are $b = 0.83 \pm 0.1$ for human A, $b = 0.71 \pm 0.07$ for human M3, and $b = 0.82 \pm 0.01$ for LFPP mice (mean ± s.d., Methods). Solid lines indicate fits to the data using maximum likelihood estimation (MLE). (d-f) Across all OTUs, the standard deviation of daily growth rates ($\sigma_\mu$) decreases with mean daily abundance ($x_m$), defined as the mean of successive log abundances, $x_m = \frac{1}{2} \log(X(t+1)) + \log \left( \langle t \rangle \right)$. Standard deviations were calculated by binning daily growth rates by different values of $x_m$ along the x-axis. Dashed lines are least-squares fits to the data, with slopes of $r = -0.16 \pm 0.02$, $-0.16 \pm 0.02$ and $-0.17 \pm 0.03$ for A, M3 and LFPP mice respectively (mean ± s.d., Methods) Growth rates in c and f were aggregated across the three mice on the LFPP diet.

In addition to species growth rates, interesting long-term dynamical trends have also been observed across different macroscopic ecosystems (Hekstra and Leibler, 2012; Keitt and Stanley, 1998; Niwa, 2007). To explore the long-term behavior of gut microbiota, we investigated how the mean-squared displacement (MSD) of OTU abundances ($\langle \delta^2(\Delta t) \rangle$) changed with time. Again, similar
to the behavior of other diverse communities (Fig. B2c), we found that the long-term dynamics of gut microbiota abundances could be well approximated by the equation of anomalous diffusion (Fig. 4.2, Fig. B3),

$$\langle \delta^2(\Delta t) \rangle \propto \Delta t^{2H} \quad (4.1)$$

where $H$ is the Hurst exponent quantifying the rate of abundance drift over time and therefore, the degree of long-term stability (Metzler et al., 2014). In contrast to normal diffusion ($H = 0.5$), a Hurst exponent of $H > 0.5$ indicates a tendency for increases (decreases) in abundances to be followed by further increases (decreases), whereas a value of $H < 0.5$ indicates a higher degree of stability and a bias for abundances to revert back to their means. Both in human and mouse gut microbiomes, our analysis revealed small Hurst exponents ($H = 0.09 \pm 0.03$, $H = 0.08 \pm 0.02$, mean $\pm$ s.d. across humans and mice). This suggests that despite overall stability (Coyte et al., 2015; Faith et al., 2013; Gibbons et al., 2017), gut microbiota exhibit a slow, continuous and predictable drift in abundances over long time periods. Furthermore, while the temporal behavior of individual OTU abundances was also well-approximated by the equation of anomalous diffusion (Fig. B4a), the distribution of Hurst exponents across individual OTUs exhibited substantial variability (Fig. B4b). This demonstrates the heterogeneity in the stability of different gut bacterial taxa within and across hosts. We show below that the stability of different taxa can be significantly affected by environmental factors such as host dietary intake.
Figure 4. 2. Long-term stability of gut microbiota abundances.

(a-c) In humans and mice, the mean-squared displacement of log OTU abundances ($\langle \delta^2(\Delta t) \rangle$) scales with time as a power law of the form $\langle \delta^2(\Delta t) \rangle \propto \Delta t^{2H}$. Hurst exponents are $H = 0.07 \pm 0.03, 0.08 \pm 0.02, 0.08 \pm 0.02$ for human A, human M3 and LFPP mice respectively (mean ± s.d., Methods). The data in c represent an average over the three individual mice on the LFPP diet (Methods). Dashed lines indicate least-squares fits to the data.

Both short and long-term dynamics of gut microbiota contribute to overall turnover in gut bacterial species. To directly investigate the dynamics of gut microbiota composition, we next calculated the distribution of residence ($t_{res}$) and return times ($t_{ret}$) for individual OTUs. Following previous macroecological analyses (Bertuzzo et al., 2011; Keitt and Stanley, 1998; Suweis et al., 2012), we defined residence times as time intervals between the emergence and subsequent disappearance of corresponding OTUs; analogously, return times were defined as the intervals between disappearance and reemergence of OTUs. Again, we observed patterns very similar to those previously described in diverse ecological communities (Bertuzzo et al., 2011; Keitt and Stanley, 1998; Suweis et al., 2012) (Fig. B2b). Specifically, the distributions of $t_{res}$ and $t_{ret}$ followed power laws, with exponential tails resulting from the finite length of the analyzed time series (Fig. 4.3, Fig. B5a,b). Notably, the distributions were also similar within and between individual human and mouse gut microbiomes ($\alpha_{res} = 2.3 \pm 0.05, \alpha_{ret} = 1.2 \pm 0.02$, mean ± s.d. across humans, $\alpha_{res} = 2.2 \pm 0.04, \alpha_{ret} = 0.72 \pm 0.03$, across mice on the LFPP diet),
suggesting that the processes governing the local emergence and disappearance of gut bacteria are likely to be independent of the specific host.

Figure 4.3. Residence and return times of gut microbiota.

(a–c) Residence \( t_{\text{res}} \) and return times \( t_{\text{ret}} \) were defined as the number of consecutive time points during which an OTU was detected at any abundance in the community or absent from the community respectively. Probability distributions for \( t_{\text{res}} \) and \( t_{\text{ret}} \) follow power laws with exponential cutoffs of the form \( p(t) \propto t^{-\alpha}e^{-\lambda t} \), with the exponential tail resulting from the finite length of each time series. Power law exponents are \( \alpha_{\text{res}} = 2.3 \pm 0.04, 2.2 \pm 0.07, 2.2 \pm 0.04 \) for residence times and \( \alpha_{\text{ret}} = 1.1 \pm 0.02, 1.2 \pm 0.05, 1.2 \pm 0.07, 1 \) for return times (mean ± s.d., humans A and M3 and LFPP mice respectively, Methods). Residence and return times are aggregated across the three individual mice on the LFPP diet. Solid lines indicate fits to the data using MLE.

Having characterized bacterial growth distributions and residence times, we next investigated the temporal variability of individual OTU abundances. One of the most general relationships in ecology that has been observed across hundreds of different biological communities is known as Taylor’s power law (Taylor, 1961; Taylor and Woiwod, 1980; Taylor et al., 1978), which connects a species’ average abundance to its temporal or spatial variance,

\[
\sigma^2_x = C \cdot \langle X \rangle^\beta
\]
where $C$ is a constant, $\langle X \rangle$ and $\sigma_X^2$ are the mean and variance of species abundances respectively, and $\beta$ is a positive scaling exponent. For processes following simple Poissonian fluctuations, the parameter $\beta = 1$, while for processes with constant per capita growth variability, $\beta = 2$ (Kilpatrick and Ives, 2003). Values of $\beta$ have been empirically observed to lie between 1 and 2 for the vast majority of investigated plant and animal species (Anderson et al., 1982). Interestingly, our analysis revealed that the temporal variability of gut microbiota also followed Taylor’s law (Fig. 4.4, Fig. B6a,b), with exponents for human and mouse gut microbiomes generally consistent with values observed previously in other ecological communities (Anderson et al., 1982) ($\beta = 1.7 \pm 0.02$ across humans, $\beta = 1.49 \pm 0.02$ across LFPP mice). This suggests that complex ecological interactions, such as competition for finite resources, may contribute to the observed species dynamics and Taylor’s law exponents in the gut microbiome (Kilpatrick and Ives, 2003). Notably, dynamics consistent with Taylor’s law have also been observed in a recent short-term analysis of the healthy human vaginal microbiome (Ma, 2015).

Although Taylor’s law described well the overall dynamics of gut microbiota, some specific OTUs clearly deviated from the general trend (Fig. 4.4). To determine whether their behavior reflected specific ecological perturbations, we identified all OTUs that exhibited significant and abrupt increases in abundance during previously documented periods of travel in human A and enteric infection in human B (David et al., 2014a) (Methods). Interestingly, these travel and infection-related OTUs corresponded to the outliers from Taylor’s law (Fig. 4.4a,b, blue circles), showing on average ~10-fold greater variance than expected based on the Taylor’s law trend (Fig. B6a,c, Table D.1). Many of these OTUs were members of the Proteobacteria (OTU 13, family: Enterobacteriaceae, OTU 29, family: Pasteurellaceae, OTU 5771, family: Enterobacteriaceae in
human A; OTU 13, family: Enterobacteriaceae in human B), which were associated with the microbiota perturbations (David et al., 2014a) (Table D.1). Moreover, other OTUs, primarily belonging to the Firmicutes, that exhibited abrupt changes in abundances at any other time point (OTU 25, family: Peptostreptococcaceae in human A; OTU 95, family: Ruminococcaceae, OTU 110, family Ruminococcaceae in human B) also displayed higher than expected temporal variability (Fig. 4.4a,b, purple circles, Fig. B6c, Table D.1). These results suggest that macroecological relationships can be used to identify and characterize specific microbial taxa that are likely involved in periods of dysbiosis and other context-specific environmental perturbations.

**Figure 4.4. Taylor’s power law in the gut microbiome.**

Mean and temporal variance of OTU abundances follow Taylor’s power law of the form $\sigma_X^2 \propto \langle X \rangle^\beta$, with $\beta = 1.66 \pm 0.09$, $1.60 \pm 0.08$, $1.49 \pm 0.02$ for humans A, B and LFPP mice respectively (mean ± s.d., see Methods). Each point corresponds to the average abundance and temporal variance of a single bacterial OTU. (a,b) OTUs that exhibited temporary and abrupt increases in abundance are indicated as colored circles (Methods). Light blue circles indicate OTUs that exhibited significant increases in abundance specifically during periods of travel (human A) and enteric infection (human B). (c) Data from each mouse on the LFPP diet are overlaid. Dashed lines indicate least-squares regression fits.

It is well established that the dynamics of diverse ecosystems are strongly affected by their environment (Brose et al., 2016). Host dietary intake is a major environmental factor influencing gut bacterial abundances (Carmody et al., 2015; David et al., 2014b; Wu et al., 2011) and disease phenotypes (Devkota et al., 2012; Turnbaugh et al., 2006). Therefore, we next explored the effects
of diet on the observed macroecological relationships describing gut microbiota dynamics. To that end, we used data from the study of Carmody et al. (Carmody et al., 2015), who investigated fecal bacterial abundances in individually-housed mice fed either a low-fat, plant polysaccharide-based (LFPP) diet, or a high-fat, high-sugar (HFHS) diet. Our analysis revealed that the short-term dynamics of gut microbiota were significantly affected by the diets. While daily growth variability declined rapidly with increasing abundance in the LFPP mice (Fig. 4.5a, green), it remained more homogeneous across OTU abundances in the HFHS mice (Fig. 4.5a, purple, regression slopes $r = -0.17 \pm 0.03$ for the LFPP diet, $-0.08 \pm 0.02$ for the HFHS diet, Z-test of regression coefficients $p= 2.0e-5$). The increased dependence of daily growth rate variability on abundance in the LFPP mice may reflect the day to day fluctuations in niche sizes on this diet. This dependence is much weaker on the HFHS diet, which may reflect the loss of differentially abundant niches on this diet, due to its significantly reduced nutrient complexity.

In addition to short-term fluctuations, we also investigated how different diets affected the long-term drift of gut microbiota. Interestingly, Hurst exponents were significantly larger in the HFHS mice, indicating substantially faster drift of bacterial abundances on this diet (Fig. 4.5b, Fig. B3b, $H = 0.19 \pm 0.02$ for the HFHS diet, $0.08 \pm 0.02$ for the LFPP diet, Z-test $p<1e-10$). Previous studies have demonstrated diet-induced compositional shifts of gut microbiota(Carmody et al., 2015; David et al., 2014b; Wu et al., 2011) and a reduced gut bacterial diversity in Western populations attributed in part to altered dietary habits(Sonnenburg and Sonnenburg, 2014; Sonnenburg et al., 2016; Yatsunenko et al., 2012). Our analysis shows that different diets not only affect the composition of gut microbiota, but also significantly change their long-term dynamics.

In addition, we found that while the abundance drift of the Bacteroidetes and Firmicutes, two major
phyla in the mouse gut, were relatively similar on the HFHS diet ($H = 0.18 \pm 0.1$ for Bacteroidetes, $H = 0.18 \pm 0.03$ for Firmicutes), the Bacteroidetes exhibited significantly reduced drift on the LFPP diet as compared to the Firmicutes ($H = 0.03 \pm 0.06$, $H = 0.09 \pm 0.02$, Z-test $p=3e^{-8}$). This suggests that while the LFPP diet decreased the long-term abundance drift of all taxa, the stability of the Bacteroidetes was particularly affected by this diet (see below).

Different diets may not only change overall gut microbiota dynamics, but also alter the temporal variability of individual taxa relative to the rest of the community. To understand taxa-specific changes, we examined Taylor’s law in mice on the LFPP and HFHS diets (Fig. 4.5c,d). This analysis showed that power law exponents were significantly different between the two diets ($\beta = 1.49 \pm 0.02$ for the LFPP diet, $\beta = 1.86 \pm 0.07$ for the HFHS diet, Z-test $p=1.5e^{-6}$). Interestingly, the temporal fluctuations of the Bacteroidetes (Fig. 4.5c,d, blue circles) exhibited significantly lower variability given their abundances on the LFPP diet, but not on the HFHS diet (hypergeometric test, $p=2.4e^{-4}$, Table D.2, Methods). Notably, Bacteroidetes are known to metabolize a wide range of dietary fibers present in the LFPP diet (Gurry et al., 2018; Martens et al., 2009; Sonnenburg et al., 2010) and are significantly lost during multigenerational propagation of mice on a low-fiber diet (Sonnenburg et al., 2016). This suggests that specific members of the Bacteroidetes (OTU 118, OTU 237, OTU 364, family: Porphyromonadaceae, Table D.2) may exhibit both lower temporal variability and abundance drift by directly exploiting stable niches that are present on the LFPP diet and likely lost on the HFHS diet. Our results also demonstrate that macroecological analyses may be used to identify specific taxa whose temporal dynamics are altered between different diets.
Figure 4.5. Dynamics of gut microbiota in mice fed different diets.

(a) OTUs in mice fed a low-fat plant-polysaccharide-based (LFPP) diet show a stronger dependence of daily growth rate variability ($\sigma_\mu$) on mean daily abundance ($x_m$) compared to those fed a high-fat high-sugar (HFHS) diet (regression slopes $r = -0.17 \pm 0.03$, $r = -0.08 \pm 0.02$; mean± s.d., LFPP and HFHS mice respectively). Data are aggregated across the three mice on each diet with dashed lines indicating least-squares regression fits. (b) OTU abundances in the LFPP mice exhibit reduced long-term abundance drift compared to those in the HFHS mice ($H = 0.08 \pm 0.02, H = 0.19 \pm 0.02$). (c,d) Taylor’s law analysis shows differences in overall scaling of average OTU abundance versus temporal variance on each diet ($\beta = 1.49 \pm 0.02, \beta = 1.86 \pm 0.07$), driven by the temporal behavior of the Bacteroidetes in the LFPP mice (blue circles). Plots correspond to data combined from the three mice on each diet. Dashed lines indicate least-squares regression performed on the combined data.
Conclusion

Despite vastly different length and interaction scales, our study reveals that global dynamical
patterns in the gut microbiome are strikingly similar to those observed in other highly diverse
ecosystems. This similarity suggests that some temporal processes in both macroscopic and
microbial communities are likely to be governed by a universal set of underlying mechanisms and
principles. Notably, a number of theoretical studies related to the neutral theory of ecology
(Hubbell, 2001) have been highly successful in capturing in other ecosystems many of the same
dynamical patterns observed in the gut microbiome (Azaele et al., 2006, 2016; Keil et al., 2010;
Pigolotti et al., 2005). Through the fundamental mechanisms of species birth, death, immigration,
random dispersal and drift, neutral theory has provided an elegant framework to understand and
connect the complex behavior of many diverse ecosystems. Despite its success, we note that many
of the assumptions of neutral theory, including its defining hallmark that differences between
species are neutral, are clearly inconsistent with the enormous taxonomic and metabolic diversity
of gut microbiota. Nevertheless, the neutral theory may be an important null model to differentiate
ecological mechanisms that are specific to the gut microbiome from those that are general to other
ecosystems.

Regardless of underlying processes, we envision that the quantitative statistical framework
developed in macroecology (Keitt and Stanley, 1998; Marquet et al., 2005) will be important for
analyzing gut microbiota dynamics. Moreover, the ability to easily perturb the composition and
environment of gut bacteria, as well as monitor their abundances at high temporal resolution,
creates an exciting opportunity to use the gut microbiome as a model system to explore general
ecological relationships. We also anticipate that a quantitative ecological framework will be useful for understanding how host-specific and environmental factors influence the dynamics of gut microbiota. Our results suggest that the observed macroecological relationships can be used to identify both global dynamical changes and also specific taxa whose abnormal temporal behavior may serve as biomarkers for periods of illness and other ecological perturbations. Therefore, to further understand the role of the gut microbiome in human health, it will be important to investigate how quantitative macroecological relationships describing microbiota dynamics vary across large and densely-sampled human cohorts.
Methods

**16S rRNA Sequence Analysis.** Raw 16S rRNA sequencing data for humans A and B was obtained from the European Nucleotide Archive (accession number: PRJEB6518(David et al., 2014a)). Raw sequencing data from humans M3, F4 and mice was obtained from the MG-RAST database(Meyer et al., 2008) (4457768.3-4459735.3 for humans; 4597621.3-4599933.3 for mice). Sequences were analyzed with USEARCH 8.1 (Edgar, 2010) using an open clustering approach. For studies including unfiltered sequencing reads, filtering was performed using the –fastq_filter command with expected errors of 2. All reads were then truncated to 100bp, with shorter reads discarded. Following a conventional approach, reads were de-replicated and clustered at 97% sequence similarity using the –cluster_otus command to generate OTUs with a minimum of 2 sequences. Sequences were then assigned to OTUs using the –usearch_global command, resulting in OTU tables for each study. Taxonomic assignments were made to OTUs using the RDP classifier (Wang et al., 2007). Sequencing reads from each sample were then rarefied to a depth of 25K, 17K and 25K for the two human studies (A/B, M3/F4) and one mouse study respectively using Qiime 1.8 (Caporaso et al., 2010).

**OTU Inclusion Criteria.** To control for technical factors such as sample preparation and sequencing noise, analysis was restricted to OTUs passing two sets of criteria. First, OTUs were required to be present in over half of the samples within respective subjects. Second, OTUs were required to have a mean relative abundance > 1e-3 over the time series. The abundance cutoff corresponded to a mean of 25 (A, B, LFPP/HFHS mice) and 17 (M3 and F4) reads over respective sampling periods. The final analysis of human individuals included ~75 OTUs comprising ~90% of the reads assigned to an OTU in any given sample. For mice, these criteria resulted in the
inclusion of ~70 OTUs in the HFHS diet and ~55 OTUs in the LFPP diet, comprising ~90% of reads assigned to an OTU in a given sample. Because the HFHS mice initially received a LFPP diet, the analysis of these mice began 5 days after the diet shift. For the calculation of residence and return times, different criteria were imposed (see below), as these analyses would be biased by a prevalence cutoff and were more robust to noise in OTU abundance levels.

**Daily growth rates.** Daily growth rates were defined as $\mu_k(t) = \log (X_k(t + 1) / X_k(t))$, where $X_k(t)$ is the relative abundance of a given OTU $k$ on day $t$. Distributions reflect community averages, with growth rates calculated for each OTU at all time points and aggregated over all OTUs. To estimate the variability of daily growth rate distributions within human subjects, each time series was divided into six consecutive time frames of equal length (estimates were insensitive to this number). Within each time frame, daily growth rates were calculated and maximum-likelihood estimation (MLE) was used to fit the Laplace distribution exponent, with the mean and standard deviation of these values reported in the main text. For the mouse study, standard deviations reflected variability across the three individual mice on each diet. Mean daily abundances $x_m$ were defined as the mean of consecutive log OTU abundances, $x_m = \frac{1}{2} [\log(X(t + 1)) + \log(X(t))]$. To estimate daily growth rate variability as a function of abundance, growth rates were binned by values of $x_m$ using a bin size of 0.4 and standard deviations $\sigma_\mu$ were then calculated on the binned growth rates. For diet comparisons, growth rates were aggregated across the three mice on each diet. Growth rates and mean daily abundances were plotted in the base ten logarithm in all figures, with the natural log used for parameter estimation.
Hurst exponents. The mean-squared-displacement (MSD) of log OTU abundances was estimated as:

$$\langle \delta^2(\Delta t) \rangle = \frac{1}{N(T - \Delta t)} \sum_k \sum_i \left[ x_k(t_i + \Delta t) - x_k(t_i) \right]^2$$

where the angled brackets denote a community average (over time and OTUs). Here, $x_k(t_i)$ is the log relative abundance of OTU $k$ at time $t_i$, $N$ is the total number of OTUs and $T$ is the total length of the time series. A maximum time lag of 100 and 15 days were chosen for human and mouse subjects respectively due to the finite length of each time series. Hurst exponents were then calculated by regressing $\langle \delta^2(\Delta t) \rangle$ against $\Delta t$ on log-transformed axes. To estimate the variability of Hurst exponents within human subjects, time series were divided into six equal-length time frames as was done for growth rate calculations. Hurst exponents for individual OTUs were estimated in a similar fashion but with displacements restricted to time averages. For diet comparisons, Hurst exponents were additionally averaged over mice within each diet:

$$\langle \delta^2(\Delta t) \rangle_{\text{diet}} = \frac{1}{L} \sum_l \frac{1}{N_l(T_l - \Delta t)} \sum_k \sum_i \left[ x_{l,k}(t_i + \Delta t) - x_{l,k}(t_i) \right]^2$$

where the outermost summation is over individual mice ($L = 3$) on each diet.

Residence and return times. Residence times ($t_{res,k}$) of an OTU $k$ corresponded to the number of consecutive time points between its appearance ($T_{a,k}$) and disappearance ($T_{d,k}$) in the community ($t_{res,k} = T_{d,k} - T_{a,k}$). Here, $T_{a,k}$ is any time point at which the OTU was detected at
a finite read count with no reads detected on the previous collection date, and $T_{d,k}$ is the next time point at which reads were no longer detected. Return times $t_{ret}$ were similarly defined as the number of consecutive time points between local disappearance ($T_{d,k}$) and reappearance ($T_{a,k}$) in the community ($t_{ret,k} = T_{a,k} - T_{d,k}$). Only intervals that fell entirely within the time frame of the study were included. A series of alternative criteria were also considered to ensure robustness of distributions. 1) To ensure results were not biased by detection sensitivity of sequencing, distributions were calculated for data subsampled to various sequencing depths (down to 1,000 reads per sample). 2) To account for false negatives in read detections, single read counts of zero interrupting a run of consecutive nonzero abundances were neglected. 3) To control for false positives, single read abundances were neglected and treated as a zero count. Results were qualitatively insensitive to both sampling depth and the alternative read detection criteria. To estimate variability of distribution parameters within human individuals, OTUs were randomized into six equal-sized groups. Residence and return times were calculated within each group and exponents were then fitted using MLE, with means and standard deviations reported in the main text. Within diets, means and standard deviations were calculated across individual mice.

**Taylor’s power law.** The mean abundance $\langle X_k \rangle$ and variance $\sigma^2_{X_k}$ for each OTU $k$ was calculated over the time series. Taylor’s law exponents were obtained by performing linear regression of the log-transformed means and variances across OTUs in each subject. To estimate variability of exponents within subjects, time series were divided into six consecutive time frames as described before. Spiking OTUs were defined as those whose abundance on any single day was greater than the average abundance over all other days by over 25-fold. Travel-related and infection-related OTUs in humans A and B were identified as those whose abundances spiked over 25-fold during
the documented time periods (David et al., 2014a). For mice, Taylor’s law outliers were identified using a likelihood-based approach. Briefly, linear regression on the log-transformed means and variances were performed on all but a single OTU $k$. The probability of observing the left out OTU $k$ was assigned using a Gaussian likelihood function based on estimated residuals. All OTUs with probability less than $\alpha = 0.025$ were taken to be outliers. For diet comparisons, means and variances were aggregated across individual mice within diets groups.

**Statistics.** All statistical analysis was performed using custom scripts written in MATLAB (https://www.mathworks.com). Comparisons of various exponents between mouse diet groups were performed by first calculating the relevant coefficient and its associated standard error using combined data across the three mice in each diet group. Z-tests were then performed by comparing the two coefficients associated with each diet group assuming normality of standard errors. Reported p-values refer to one-sided tests.
References


David, L.A., Materna, A.C., Friedman, J., Campos-Baptista, M.I., Blackburn, M.C., Perrotta, A.,


CHAPTER 5: LIPIDS ARE A MAJOR GROWTH LIMITATION IN CELLS DEFICIENT FOR ELECTRON ACCEPTORS

Author statement

Some passages and figures have been adapted or quoted verbatim from the article:
Lipids are a major growth limitation in cells deficient for electron acceptors, 2018. In Preparation.

Authors

Brian W. Ji\textsuperscript{1}, Zhaoqi Li\textsuperscript{2}, Purushottam D. Dixit\textsuperscript{1}, Lucas B. Sullivan\textsuperscript{2}, Matthew G. Vander Heiden\textsuperscript{2*} & Dennis Vitkup\textsuperscript{1,3*}

Author affiliations

\textsuperscript{1}Department of Systems Biology, Columbia University, New York, NY, USA. \textsuperscript{2}The Koch Institute for Integrative Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. \textsuperscript{3}Department of Biomedical Informatics, Columbia University, New York, NY, USA.
Mammalian cell proliferation requires the availability of terminal electron acceptors such as oxygen. However, the downstream anabolic processes supported by exogenous electron acceptor availability and their quantitative relationships to cell growth are not well understood. Here, we use a joint computational and experimental approach to identify lipid production as the key determinant of cellular growth rate under electron acceptor limitation. We find that cells relying on de novo lipogenesis are highly sensitive to mitochondrial electron transport chain (ETC) inhibition. This sensitivity results from the substantial requirements of NAD+ for lipid biosynthesis, which cannot be sufficiently regenerated from mitochondrial complex I when respiration is inhibited. Providing cells with exogenous pyruvate, an alternative electron acceptor to oxygen, rescues cell growth independently of its role as a carbon or ATP source. Instead, pyruvate restores intracellular NAD+ levels, thereby increasing de novo synthesis rates of the major membranes lipid species, and as a consequence, restoring cellular proliferation. Importantly, bypassing the NAD+ requirements for de novo lipogenesis by directly providing cells with exogenous lipids or acetate rescues growth sensitivity to ETC inhibition in a dose-dependent fashion. These results suggest that the ability to produce or acquire lipids is a major growth determinant of cells deficient for electron acceptors, opening up therapeutic avenues to target tumors in vivo.
Results

In proliferating mammalian cells, biomass synthesis often requires a series of oxidative reactions involving the transfer of electrons to the cofactor nicotinamide adenine dinucleotide (NAD+) to generate its reduced form, NADH (Hosios and Vander Heiden, 2018). To sustain proliferation, NADH must ultimately transfer these electrons to a terminal electron acceptor that is then excreted from the cell, thereby regenerating NAD+ and allowing the cycle to continue (Fig 5.1a). In respiration-competent cells, molecular oxygen fulfills this role through the activity of the mitochondrial electron transport chain (ETC), but other compounds such as pyruvate may supplement and even taken the place of oxygen as a terminal electron acceptor (King and Attardi, 1989; Sullivan et al., 2015).

We and others have previously demonstrated that an important requirement downstream of respiration-mediated NAD+ regeneration is the synthesis of the amino acid aspartate (Birsoy et al., 2015; Sullivan et al., 2015). We reasoned that, in addition to aspartate, the synthesis of other biomass components would also require NAD+. Therefore, using a genome scale model of human cell metabolism (Duarte et al., 2007), we calculated the NAD+ requirements for the de novo synthesis of all major cellular biomass components using canonical or established routes from literature (Methods). Surprisingly, this analysis revealed that by far the largest demand for NAD+ was for the synthesis of membrane phospholipids and sterols (Fig. 5.1b). While lipids are paradoxically known to be reduced overall relative to most extracellular nutrients (Hosios and Vander Heiden, 2018), our analysis identified the initial oxidative reactions converting extracellular nutrients to the lipogenic precursor acetyl-coA to require substantial amounts of
NAD+. This was true in the model regardless of whether acetyl-coA was synthesized from glucose or glutamine (Metallo et al., 2011; Mullen et al., 2012) (Methods).

Figure 5. 1. Lipids biosynthesis represents a major demand for NAD+ and confers sensitivity of cancer cells to mitochondrial inhibition.

(a) Schematic of NAD+ and NADH production and consumption fluxes in relation to cellular growth. The production of biomass from extracellular nutrients generates NADH, which must be converted back to NAD+ through donation of electrons to a terminal electron acceptor. Different cellular biomass components (lipids, amino acids, and nucleotides) required differing amounts of NAD+ for their synthesis from extracellular nutrients. At steady state, the consumption rates of terminal electron acceptors (molecular oxygen) must match the NAD+ requirements to sustain a particular proliferation rate. (b) Minimal requirements of NAD+ regeneration for the synthesis of various biomass components. Estimates were calculated using a genome-scale model of human metabolism. Numbers represent the amount of NAD+ required to synthesize the indicate biomass components per unit growth rate. (c) Model of cellular growth under electron acceptor limitation. Relative rates of de novo lipid synthesis determine the growth rate through the diversion of NAD+ away from other biomass components. (d) Proliferation rates of HeLa and H1299 cells in media supplemented with or deprived of exogenous lipid sources. (e-g) Proliferation rates of HeLa and H1299 cells in media supplemented with or deprived of exogenous lipid sources in the presence of the electron transport chain inhibitors phenformin (complex 1), rotenone (complex 1) and antimycin (complex III). Error bars denote the standard deviation over biological triplicate measurements. *p<0.05 one-sided test.
Given the significantly higher NAD+ requirements associated with lipid biosynthesis as compared to other biomass components, we hypothesized that lipid synthesis rates were the primary growth determinant of cells limited for electron acceptors. Specifically, cells relying entirely on de novo lipid synthesis would be highly sensitive to electron acceptor limitation. This is because, for a given rate of oxygen consumption and NAD+ regeneration through the mitochondrial ETC, a vast majority of the available NAD+ would be subsequently consumed by lipid synthesis reactions (Fig. 5.1c). A much smaller fraction of the NAD+ would then be available for the synthesis of other biomass components such as amino acids and nucleotides, resulting in a slower proliferation rate. On the other hand, for cells capable of exogenous lipid uptake that required very little de novo lipid synthesis, the same supply of NAD+ would support a faster proliferation rate by enabling an increased flux through the remaining oxidative, anabolic reactions (Fig. 5.1c).

To test this prediction, we first grew HeLa and H1299 cancer cells in media replete with or depleted of exogenous lipid sources (Methods, Fig. D.1). We observed similar proliferation rates of cells in the presence or absence of lipids, indicating that cancer cells can rely solely on de novo lipogenesis when lipids are not present in the environment (Fig. 5.1d). Consistently, we found that rates of palmitate synthesis in HeLa cells were significantly elevated in delipidated conditions (Fig. D2). Given the observed differences in lipid synthesis rates, we predicted that cells deprived of exogenous lipids would then require larger rates of NAD+ regeneration to support de novo lipogenesis. Indeed, we observed significantly increased oxygen consumption rates of HeLa cells grown in delipidated conditions as well a significantly increased intracellular NAD+/NADH ratio (Fig. D3). The increased oxygen consumption rate and NAD+/NADH ratio corresponded to increased citrate pool sizes, the major precursor for de novo fatty acid synthesis (Fig. D3), as well
as an increased flux through pyruvate dehydrogenase, the primary route of cytosolic citrate production under normoxic conditions. Thus, increased lipids synthesis rates are associated with increased oxygen consumption rates and intracellular NAD+ to support flux through the lipid synthesis pathway, as predicted by the model.

To limit rates of oxygen consumption and NAD+ regeneration, we treated cells with a moderate dose of the biguanide phenformin, an inhibitor of mitochondrial complex I. Interestingly, low doses of phenformin significantly reduced the proliferation of both HeLa and H1299 cells in lipid-depleted conditions while having only modest effects on cells grown in the presence of lipids. (Fig. 5.1e) To determine whether this differential growth sensitivity was specific to phenformin, we treated cells with the mitochondrial ETC inhibitors rotenone and antimycin, targeting complexes I and III respectively. We found that for both inhibitors, cell proliferation in lipid-depleted conditions was significantly attenuated compared to lipid-replete conditions, suggesting that the decreased proliferation was related to impaired mitochondrial respiration (Fig. 5.1f,g).

We hypothesized that the observed sensitivity of cells to ETC inhibition was a due to insufficient NAD+ regeneration to sustain rates of de novo lipogenesis when lipids were depleted from the environment. Consistent with this hypothesis, both citrate pool sizes and flux through PDH were significantly attenuated upon treatment of cells to phenformin (Fig. D3). In the absence of functional mitochondrial respiration, alternative electron acceptors to oxygen, such as pyruvate, may continue to support proliferation (King and Attardi, 1989; Sullivan et al., 2015). We therefore hypothesized that supplying exogenous pyruvate would restore proliferation by regenerating NAD+ through cytosolic lactate dehydrogenase, enabling increased flux through de novo lipid
synthesis pathways. Indeed, the addition of pyruvate significantly increased the proliferation of HeLa and H1299 cells treated with phenformin in lipid-depleted conditions (Fig. 5.2a,b). Consistent with its role as an alternative electron acceptor to oxygen, pyruvate also significantly increased the intracellular NAD+/NADH ratio (Fig. 5.2c,d). To ensure that pyruvate was not acting as a source of carbon or ATP to cells, we sought a way to increase intracellular pyruvate levels without increasing NAD+ regeneration rates. Given the reversible nature of cytosolic LDH, we reasoned that supplying cells with lactate would decrease the consumption of pyruvate through LDH, thereby increasing intracellular pyruvate availability while at the same time reducing NAD+ regeneration. We found that supplementing cells with lactate did not increase cell proliferation or intracellular NAD+/NADH levels, suggesting that the pyruvate-mediated growth rescue was due to its role as an exogenous electron acceptor in this context (Fig. 5.2a,b).
Figure 5.2. Pyruvate the rescues growth sensitivity of cells to ETC inhibition by supporting de novo lipogenesis.

(a,b) Proliferation rates of phenformin-treated HeLa and H1299 cells in media depleted of exogenous lipids and supplemented with either pyruvate or lactate. (c,d) Intracellular NAD+/NADH ratios of HeLa and H1299 cells in the same conditions a and b. Error bars denote the standard deviation of triplicate measurements. *p < 0.05 one-sided t-test.

We next considered why cells growing in lipid-replete conditions were relatively resistant to the effect of phenformin. If metabolic limitation for lipids was downstream of a limitation for electron acceptors, we reasoned that lipid production routes that bypassed reactions consuming NAD+ could also rescue the growth sensitivity of cells to ETC inhibition. The most direct route for cells to acquire lipid membrane species is their uptake from the extracellular environment. We therefore
supplemented lipid-deprived cells with a mixture of extracellular phospholipids and sterols. Strikingly, the uptake of exogenous lipids was associated with a significant increase in proliferation of phenformin-treated cells, establishing lipids as the primary downstream metabolic requirement fulfilled by electron acceptor availability. In addition to exogenous lipids, extracellular acetate can also be an important carbon source of lipogenesis through the activity of coA-synthetase 2 (ACSS2) (Schug et al., 2015). Importantly, the synthesis of lipid biomass precursors from acetate bypasses any oxidative reactions that consume NAD+. We found that providing cells with exogenous acetate also significantly rescued the growth of cells in the presence of phenformin (Fig. 5.3a,c). Consistent with its role as a direct carbon source for lipid biosynthesis, acetate did not alter the intracellular NAD+/NADH ratio (Fig. 5.3b,d). Expectedly, neither acetate nor the lipid mixture increased the proliferation of phenformin-treated cells, when lipids were already available in the environment (Fig. 5.3a,c). Collectively, our results support a model in which lipids represent a major metabolic limitation of cells deficient for electron acceptors. Cell proliferation can be restored by increasing the availability of exogenous electron acceptors to support de novo lipogenesis, or providing alternative sources of membrane lipids that bypass the requirement of electron acceptors.
Figure 5.3. Lipids are the primary metabolic limitation of cancer cells limited for electron acceptors.

(a,c) Proliferation rates of HeLa and H1299 cancer cells across indicated conditions. (b,d) Intracellular NAD+/NADH ratios in HeLa and H1299 cancer cells across incided conditions. (e) Correlation of a consensus hypoxic gene expression signature with the expression of 87 KEGG metabolic pathways (Methods). Fatty acid biosynthesis is the most negatively correlated pathway with hypoxia. (f) Correlation between hypoxic gene expression score and fatty acid biosynthesis score across over 10,000 primary tumors obtained from TCGA.
Our finding in cell culture raise the possibility that tumors in vivo have evolved mechanisms to sustain proliferation in conditions of electron acceptor limitation, such as hypoxia or pseudohypoxia (Gomes et al., 2013). We therefore investigated whether the mRNA expression levels of certain metabolic pathways were correlated with previously-defined expression signatures of tumor hypoxia (Harris, 2002; Mense et al., 2006). Strikingly, when we analyzed RNA-seq data from over 10,000 different TCGA tumor samples, we found lipid biosynthesis to be the most negatively correlated pathway to the tumor hypoxia expression signature (Fig. 3e) (Pearson’s R = -0.38, p < 1e-10). This result could not be explained by decreased tumor growth, as a significant correlation was observed even after controlling for the expression of other pathways strongly associated with tumor growth, such as purine and pyrimidine metabolism and amino-acyl tRNA biosynthesis (partial correlation R = -0.38,-0.38, -0.33 respectively).

Furthermore, the expression of CD36, the major fatty acid membrane transporter, was positively correlated with hypoxia across all primary tumors (Pearson’s R = 0.33), as was ACCS2 in a subset of tumor types. These results suggest that tumors in vivo cope with conditions of electron acceptor limitation through the downregulation of de novo lipid synthesis and the concomitant upregulation of alternative lipid acquisition strategies. Therefore, the growth limitation we identify in cell culture appears to be highly relevant in primary tumors as well.

**Discussion**

Here, we have used quantitative modeling to understand the metabolic constraints of cell growth under electron acceptor limitation. Our study identifies the ability of cells to produce or acquire lipids as the primary determinant of cell growth in conditions when electron acceptors are limiting.
This provides a simple and unifying framework to understand seemingly disparate lipid acquisition strategies of cancer cells in vitro and in vivo.

It is well-established that in response to hypoxia, cancer cells increase rates of lipid uptake from the extracellular environment. In addition to protection upon re-oxygenation (Bensaad et al., 2014) or the bypass of any reactions directly consuming oxygen (Kamphorst et al., 2013), our study supports an additional and more fundamental role for lipid uptake: to enable lipid biomass production when electron acceptors become limiting. Consistent with this idea, it has also been found that tumors upregulate expression of ACCS2 in conditions of hypoxia (Schug et al., 2015). Our results suggest that direct uptake of lipids and the utilization of acetate provide alternative, less oxidative routes of lipid membrane synthesis that bypass the significant requirements of NAD+ associated with the more canonical routes involving glucose and glutamine. Together with our analysis of primary tumor expression, it seems that tumors have evolved complex strategies to support lipid membrane biosynthesis and sustain proliferation in the harsh conditions of the tumor microenvironment, where electron acceptors such oxygen are likely to be in short supply.

Our results also provide an explanation for the preferential use of glutamine as a carbon source for lipid biosynthesis in conditions of hypoxia or mitochondrial dysfunction (Metallo et al., 2011; Mullen et al., 2012). De novo lipogenesis typically involves the entry of glucose-derived pyruvate into the TCA cycle and the subsequent efflux of citrate to produce cytosolic acetyl-coA. It has been suggested that glutamine carbons may supplement the intracellular citrate pool when glucose carbons are diverted away from the TCA cycle in hypoxia or mitochondrial dysfunction. We propose that underlying this behavior is a growth strategy when electron acceptors, not lipid
carbons, are limiting. Following our model calculations, the production of acetyl-coA from glucose requires ~2 molecules of NAD+ per molecule of acetyl-coA synthesized, while the route from glutamine requires at most 1 molecule of NAD+ per molecule of acetyl-coA. Therefore, when the regeneration of NAD+ is limiting for growth, proliferation can be increased using a strategy involving glutamine as opposed to glucose.

Our study paints a more complex and nuanced picture of the metabolic limitations of cancer cells. While we demonstrate that lipids are a major growth limitation in cells depleted for electron acceptors, they are not the only metabolic limitation. In conditions in which lipid biomass requirements can be completely fulfilled through exogenous uptake, other metabolites can then become growth limiting. Indeed, calculations from the genome-scale model of human metabolism predicted that a significant NAD+ demand was for the synthesis of aspartate. Therefore, in conditions where extracellular lipids are in excess, aspartate likely becomes the primary metabolic limitation of proliferating cells, a finding consistent with previous studies (Birsoy et al., 2014; Sullivan et al., 2015).

Finally, our results identify a potential vulnerability of tumor growth in vivo. Future studies are required to determine the efficacy of strategies to either target the uptake of lipids from the environment, or reduce the extracellular lipid content. Such strategies, in combination with electron transport chain inhibitors such as metformin, may hold particular promise to target the growth of some tumors.
Computational Methods

Estimation of NAD+ costs associated with the synthesis of various biomass components.

Generation of a cancer cell-specific metabolic network

We began with a genome-scale FBA model of human cellular metabolism (Duarte), containing a total 3,744 reactions and 2,766 metabolites. In order to remove any reactions that were unlikely to carry flux in cancer cells and could potentially confound downstream analysis, we developed a custom model-pruning procedure. Specifically, we calculated the average growth rate and average consumption and excretion rates of all major nutrients measured across the NCI-60 panel of cancer cell lines (Jain). We then constrained the full genome-scale FBA model to these measured values to obtain a representative, consensus FBA model of cancer cell metabolism. To remove reactions that were unlikely to carry flux in our network, we randomly selected reactions beginning with the set of 3,744 original reactions and asked whether the removal of that reaction would result in a model that was incompatible with the constrained average measured consumption and excretion rates and growth rate. If the model became infeasible without this reaction in the network, the reaction was kept. If not, it was removed. This process was repeated until no further reactions could be removed without generating an infeasible model. This procedure resulted in a single minimal model consistent with experimental measurements. However, this minimal model was not unique, as it depended on the order in which reactions were chosen and removed from the original network. We repeated the entire process 1,000 times to generate 1,000 minimal consistent models. For each of the 3,744 reactions in the original network, we then calculated the fraction of times that reaction was observed in any of the minimal networks, representing a probability that a
reaction was required to carry flux to produce originally measured exchange rates and growth rate. Using these calculated probabilities, we searched for the maximum probability for which all reactions with probabilities above this value would be sufficient generate a metabolic network that was compatible with experimental data. The resulting metabolic network contained 600 reactions and was used for all downstream analysis.

*Calculating the NAD+ requirements of lipid biosynthesis*

To calculate the amount of NAD+ required for de novo lipid synthesis, we utilized the biomass production reaction in the pruned FBA network from the previous section. This reaction contains the stoichiometry of cellular biomass composition for an average mammalian cell. In order to calculate the lipid-specific NAD+ requirements, we modified the biomass production reaction to include only lipid metabolite species (Table D1), while preserving their stoichiometry in biomass. We set the flux through the biomass to equal to one (units of 1/hr) and solved for a flux distribution that minimized the total amount of NAD+ generated in both the mitochondria and cytosol. This value was taken to represent the minimum NAD+ required per unit growth rate for the synthesis of all membrane lipid species. To calculate the glucose-specific NAD+ costs of lipid synthesis, we solved the FBA model while limiting glucose to be the sole organic carbon source in the model. To calculate the NAD+ costs of lipid biosynthesis from glutamine, we allowed both glucose and glutamine as organic carbon sources. The resulting flux distribution utilized the reductive carboxylation pathway as the sole source of lipid carbons. We therefore took this to reflect the minimum amount of NAD+ required for lipid synthesis from glutamine.
Calculating the NAD+ requirements of aspartate, serine and nucleotide biosynthesis

We used a similar approach as described above to calculate the NAD+ requirements for the synthesis of other biomass species. To calculate the NAD+ costs of aspartate synthesis, we removed all metabolic species in the FBA biomass production reaction except for the aspartate contribution to the cellular proteome. Therefore, the calculated NAD+ cost reflects the aspartate requirement for protein synthesis but does not include the NAD+ generated for the role of aspartate in nucleotide synthesis (this cost is incorporated in the nucleotide synthesis estimation below). As studies have demonstrated that aspartate is poorly transported across the cell membrane and is typically synthesized using glutamine carbons, we limited glutamine to be the only source of organic carbon in the FBA model. An identical approach was used for the estimation of NAD+ production required for serine synthesis, with the exception of glucose being the sole carbon source instead of glutamine. Again, the NAD+ costs associated with serine synthesis reflect only the demand for serine in protein synthesis, and does not include the requirements for serine in lipid and nucleotide biosynthesis (these costs are instead incorporated in the lipid and nucleotide NAD+ requirements).

To calculate the NAD+ requirements associated with de novo nucleotide synthesis, we restricted the FBA biomass reaction to DNA and RNA nucleotides. We then solved for the minimal NAD+ required for the their collective synthesis at a growth rate of one, while limiting organic carbon from glucose and glutamine. Notably, the nucleotide-associated NAD+ cost includes the requirement for serine carbons to fuel one carbon metabolism, as well as the contributions of aspartate in both pyrimidine and purine biosynthesis.
RNA expression analysis of primary tumors. mRNA-seq data from roughly 11,000 primary
tumors spanning 32 tissue types was obtained TCGA. Metabolic gene annotations were obtained
from the KEGG data base (cite). Expression scores for each metabolic pathway $i$ in tumor sample
$j$ were defined as:

$$\text{Score}_{ij} = \langle \log (X_{kj}) \rangle_{k \in \text{pathway } i} - \langle \log (X_{mj}) \rangle_{m \in \text{all genes}} \quad (5.1)$$

where $X_{kj}$ corresponds to the expression level (units of TPM) of gene $k$ in tumors sample $j$. The
first term on the right-hand side of equation (1) is the average log expression of all genes within
a given pathway $i$. The second term reflects an average over all the genes in the tumor sample.
To calculate hypoxia scores for each tumor sample, we first generated a consensus hypoxia gene
signature set by merging three independent hypoxia gene sets obtained from the Molecular
Signatures Database (cite). Any KEGG metabolic genes were removed from the hypoxia gene
set and hypoxia scores were calculated following equation (5.1).
References


Mullen, A.R., Wheaton, W.W., Jin, E.S., Chen, P.H., Sullivan, L.B., Cheng, T., Yang, Y., Linehan,


CHAPTER 6: CONCLUSION

Summary

Here, we have developed top-down quantitative approaches to study both the spatiotemporal dynamics of human gut microbiota and the metabolic limitations of cancer cellular growth. We first show that an inherent limitation of current microbiome sequencing studies is the effect of different sources of variability on measured microbiota abundances. These sources of variability prohibit the analysis and interpretation of longitudinal time series studies of the gut microbiome or any other microbial community. Saliently, we demonstrate that the observed dynamics of over half of the detected bacterial taxa in both the human gut and a soil community in Central Park can be primarily attributed to technical noise. Furthermore, we also demonstrate significant spatial heterogeneity of fecal microbiota, which is often neglected in gut microbiome studies. We develop a joint computational and experimental framework, DIVERS, that provides a mathematically principled solution to quantify the sources of variability in microbiome sequencing studies, measured in absolute abundances. Furthermore, our method quantifies the major sources of covariation of pairs of microbial taxa in the human gut. Importantly, DIVERS requires minimal sample collection, preparation and sequencing as compared to naïve protocols, facilitating its integration into existing sequencing pipelines.

We next investigate the ecological dynamics of human and mouse gut microbiota. Borrowing ideas from macroecology, we describe global aspects of gut microbiota dynamics, including the nature of daily abundance fluctuations, long-term drift, residence and return times and the scaling of mean
taxa abundance with temporal variance. We show that the statistical relationships describing gut microbiota abundances are highly similar to other biological and inanimate systems, suggesting that universal mechanisms may structure the dynamics of gut bacteria. Finally, we demonstrate that the observed scaling relationships in the gut microbiome enable identification of taxa exhibiting altered dynamics in a context-specific fashion. These include individual species that experience surges in abundance upon travel or enteric infections in humans, as well as global dynamical changes associated with dietary perturbation in mice. These results provide a collective framework for investigating the dynamics of human and mouse gut microbiota across different environmental contexts and conditions.

Finally, we employ coarse-grained mathematical modeling to identify a major metabolic limitation of cancer cells. By calculating the redox requirements associated with cell growth using a genome-scale metabolic model of human metabolism, we show computationally that lipid synthesis is a major oxidative burden to cells with finite capacity for oxygen uptake and NAD+ regeneration. As a consequence, we predicted that lipid biosynthesis rates cells would be the primary growth determinant of cells deficient for electron acceptors. Our predictions were confirmed experimentally, identifying lipids as a major metabolic limitation of cancer cell growth in vitro. In vivo, we identify the lipid biosynthesis pathway as being highly downregulated at the mRNA level in primary tumors experiencing hypoxia, with alternative lipid acquisition strategies such as direct extracellular lipid uptake or synthesis from acetate being significantly upregulated in certain tumor types. Collectively, our results highlight the vulnerability of cancer cells undergoing de novo lipid synthesis to electron acceptor limitation, with implications for targeting tumor growth in vivo.
DIVERS: A generalized dual-reporter method for sequencing studies

The challenge of understanding the sources of variability in biological systems is ubiquitous to nearly all fields of biology. A central challenge in systems biology has to identify the factors contributing to variability in gene expression levels across a population of genetically identical cells, and how this variability may influence phenotypic heterogeneity, the functioning of gene networks, and long-term evolutionary strategies (Balaban et al., 2004; Kussell and Leibler, 2005; Ozbudak et al., 2002; Raser and O’Shea, 2004). A fundamental question is the degree to which the gene expression process is inherently noisy within individual cells (i.e. whether noise is primarily intrinsic to cells), or whether cell to cell differences across a population contribute to variability in gene expression (i.e. whether noise is primarily extrinsic to cells) (Hilfinger and Paulsson, 2011). In 2002, Elowitz and colleagues published a seminal paper describing a dual-reporter method to separately quantify the contributions of intrinsic or extrinsic noise to gene expression profiles of single cells (Elowitz et al., 2002; Swain et al., 2002). The method involved expressing two identical, fluorescent protein copies (YFP and CFP) within individual cells and performing computations on these expression levels across the cell population. The simplicity and elegance of the dual-reporter method, in conjunction with other early seminal work, has spearheaded nearly two decades of investigation into the causes and consequences stochastic gene expression in both bacteria and eukaryotes (Bar-Even et al., 2006; Cai et al., 2006; Newman et al., 2006; Sigal et al., 2006; Taniguchi et al., 2011).

DIVERS is, at its core, a generalized dual-reporter approach for microbiome studies. Notably, replicate sampling from two spatial locations of a single stool specimen to separate temporal from
spatial sources of variability of microbiota abundances is directly and conceptually analogous to expressing two identical fluorescent protein copies within a single cell to separate extrinsic from intrinsic sources of noise in gene expression levels. The additional wrinkle of DIVERS is the decomposition of a third variance contribution, representing technical variability in sequencing studies. Importantly, while the underlying mathematical framework of DIVERS is constant, the specific biological contexts to which DIVERS can be applied may vary drastically. Therefore, it is our hope that DIVERS may represent a generalized dual reporter method for all types of sequencing studies. For example, given the increasing interest in capturing tumor expression heterogeneity up to the single cell level, DIVERS may be used to understand patterns of expression variability across different physical locations of tumors, as well as across specific cell types and subpopulations (Patel et al., 2014; Puram et al., 2017; Tirosh et al., 2016). Furthermore, with the recent advancements in the sequencing capabilities of human immune-cell repertoires, DIVERS may also be applied to monitor cellular immune responses over time, and their variability across different tissue compartments (Wargo et al., 2016). Beyond sequencing studies, a recent dual-reporter approach was used to understand patient-specific glucose responses to standardized meals, revealing significant inter-individual heterogeneity in post-prandial glucose levels (Zeevi et al., 2015). These and other examples highlight the generality of the DIVERS approach.

Towards the investigation of human gut microbiota dynamics across populations

Measurements of static gut bacterial abundances have been instrumental to our current understanding of the human gut microbiome in health and disease. Analysis of population-wide
cohorts have identified individual taxa and gene categories enriched or depleted in a number of human diseases (Le Chatelier et al., 2013; Frank et al., 2007; Vandeputte et al., 2017; Wang et al., 2012). However, many processes remain undetectable when simply analyzing average abundance changes across conditions (David et al., 2014; Wu et al., 2011). In contrast, we identify rich, dynamical information from high-resolution time series profiling of healthy human individuals that can be leveraged in concert with tractable quantitative frameworks to yield novel insights into the temporal processes governing human gut microbial ecology. For example, our analysis of human gut microbiome dynamics using a macroecological approach identified specific taxa associated with travel and diarrheal events, as well as enteric infection. We also demonstrate global alterations in dynamics mice fed a low-fat, plant diet compared to a diet high in simple sugar and fat content. While analysis of average abundances between diet groups revealed an overall increase in the relative abundance of the Bacteroidetes on the LFPP diet (Carmody et al., 2015), analysis of dynamical behavior revealed that this increase in abundance was associated with particularly low variability, suggesting that the Bacteroidetes increase their abundance by exploiting highly stable niches on the LFPP diet. Consistently, we observed marked long-term stability of these same Bacteroidetes OTUs. With the increasing scale and decreasing costs of metagenomic sequencing studies, we hope that the presented macroecological analyses of the healthy human gut microbiome will inspire high-resolution time series profiling of larger healthy populations as well as diseased cohorts. Such an approach will be particularly important in diseases for which a single bacterial taxa has not been causally identified (Ley et al., 2005; Vatanen et al., 2018). The additional dimension revealed by dynamical measurements may shed insight into the underlying ecological mechanisms that are perturbed in the dysbiotic human gut.
Coarse-grained approaches for studying cancer cell physiology

The heterogeneity of tumor metabolism has been well studied, representing a major therapeutic challenge. Tumor metabolic phenotypes may be a complex function of genetic background, lineage effects and the surrounding extracellular environment (Vander Heiden and DeBerardinis, 2017). This has led to seemingly disparate therapeutic strategies across genetically and histologically distinct tumor types (Davidson et al., 2016; Mayers et al., 2014). However, there exist organizing biological principles that define the growth of all mammalian cells. The liability of lipid synthesis under electron acceptor limitation is one outcome of taking a more coarse-grained view of cancer cell metabolism and physiology. Because cell growth requires, at the quantitative level, the production of various biomass components at distinct stoichiometries defined by the cellular biomass composition, simple principles such a mass and flux balance provide a useful quantitative framework to rationalize existing observations and predict new vulnerabilities. For example, cancer cells take up large amounts glucose, while at the same time carrying out high rates of oxidative phosphorylation (Fan et al., 2013). Surprisingly, through simple mathematical calculations, the rate of ATP production through glycolysis and oxidative phosphorylation corresponds exactly to the amount of ATP required for cellular translation at a given growth rate (unpublished data), suggesting that the majority of cellular ATP drives protein translation. Interestingly, if the rate of glucose intake is reduced through modulation of extracellular glucose concentration, the rate of oxygen consumption is increased such that the total ATP production remains constant. This simple example suggests that rapid rates of glucose uptake are not required for ATP production, because the cell may maintain an equivalent growth rate through the upregulation of oxidative phosphorylation. Thus, the Warburg effect observed in
cancer cells cannot be attributed to increased rates of ATP production. While this coarse-grained analysis may enable identification of general growth principles of cancer cells, it necessarily sacrifices molecular mechanisms and details.

A coarse-grained approach is not specific to understanding metabolic strategies of cancer cells. In bacterial and yeast cells, it is well known that the ribosomal content increases linearly with the growth rate (Scott et al., 2010). This rather surprising relationship is a simple result of cells balancing catabolic rates of extracellular nutrients into amino acids with rates of incorporation of those amino acids into proteins. Because flux balance dictates that amino acid be produced and consumed at the exact same rate (i.e. the cell growth rate), a series of constraint exists on the activity of other intracellular processes. For example, regulation of translation through the well-characterized mTOR pathway must be consistent with observed translation rates. This further constrain patterns of ribosomal occupancy in the cell, which must also be consistent with other intracellular processes such as rates of transcription, and protein degradation (Ingolia et al., 2011; Lalanne et al., 2018; Li et al., 2014). Thus, quantitative relationships between these various intracellular processes constrains that set of possible intracellular states compatible with an observed phenotype (i.e. cellular growth rate). Knowledge of such constraints restricts the set of underlying mechanisms giving rise to the phenotype in question. While such approaches have yielded novel insights into bacterial systems, they remain vastly underexplored in cancer cells.
References


APPENDIX A: SUPPLEMENTARY FIGURES AND TABLES FOR CHAPTER 3

Author statement

Some figures and tables have been adapted from the article:

Authors

Brian W. Ji$^{1,2†}$, Ravi U. Sheth$^{1,2†}$, Purushottam D. Dixit$^1$, Andrew Kaufman$^1$, Harris H. Wang$^{1,3*}$ & Dennis Vitkup$^{1,4*}$

Author Affiliations

$^1$Department of Systems Biology, Columbia University, New York, NY, USA.
$^2$Integrated Program in Cellular, Molecular, and Biomedical Studies, Columbia University, New York, NY, USA.
$^3$Department of Pathology and Cell Biology, Columbia University, New York, NY, USA
$^4$Department of Biomedical Informatics, Columbia University, New York, NY, USA.
†These authors contributed equally to this work.
Figure A 1. Assessment of the DIVERS variance decomposition model using simulated bacterial dynamics.

(a) Stochastic model of spatiotemporal dynamics in an interacting bacterial community. Simulations were performed on a 10 x 10 lattice with continuous boundary conditions, where dynamics of each species were governed by birth events, death events or random migration to neighboring locations. Species interactions were modeled using density-dependent logistic growth at each location (Online Methods). Results for four species with non-zero steady state abundances are shown. (b) Expected contributions of temporal, spatial sampling and technical sources to total abundance variance, and predictions from the DIVERS variance decomposition model. Technical variability was modeled using Poisson sampling noise centered on the true abundances at each spatial location (Online Methods). Results for four species in the simulated community are shown and correspond to those in panel a. Expected variance contributions were empirically calculated from simulated data (Online Methods). Both expected and DIVERS variance contributions were calculated across twenty time points after any initial transient behavior. Error bars represent standard deviations based on 10,000 re-samplings from different pairs of spatial locations in the environment.
Figure A1. Validation of the spike-in sequencing approach to estimate total bacterial loads in collected fecal and soil samples.

(a) Relative endogenous abundances of the spike-in strain, Sporosarcina Pasteurii, in eighteen fecal samples collected from the same human individual presented in the main text. No reads belonging to the spike-in strain were detected in any samples. (b) Relative endogenous abundances of Sporosarcina Pasteurii across twenty of the soil sampling sites in Central Park. Observed relative abundances are consistent with sample read-through (contamination from other samples in the same sequencing run) and are nearly two orders of magnitude lower than sequenced abundances in the same samples for which the spike-in strain was added. (c) Expected and observed behavior of spike-in strain abundances across two serially diluted fecal samples (Online Methods). Expected behavior was derived based on the dilution factor of 2 used in the dilution series of each original fecal sample (Online Methods). (d) Technical replicate measurements of total bacterial densities in fecal and soil samples. Measurements are separately normalized to a mean of one within fecal or soil samples. Technical replicates are highly correlated (Pearson’s r = 0.9).
Figure A 2. Variance decomposition of OTU relative abundances in the human gut microbiome.

OTUs are binned by mean relative abundance across samples. Stacked bars indicate the average fraction of total variance attributed to temporal, spatial sampling and technical sources for OTUs within each bin. Error bars denote the SEM.
Figure A 3. Variance decomposition for high and low-abundance OTUs in the human gut microbiome.

Temporal, spatial sampling and technical contributions to total variance for all OTUs with (a) mean absolute abundance > $10^{-4}$ and (b) mean absolute abundance < $10^{-4}$. Boxes show the median and interquartile ranges, with maximum whisker lengths three times the interquartile range. (c) OTU abundance variability due to technical noise. Multiple technical replicates (n=12) were processed from fecal samples obtained from a single spatial location of a stool specimen. Purple dots show the normalized technical variability (variance/mean$^2$) as a function of average abundance across twelve technical replicates. Technical noise profiles obtained from DIVERS are shown in gray dots. The inverse scaling expected from Poissonian sampling noise is indicated with the dashed line with slope = -1. A noise floor is observed at high OTU abundances (indicated by the horizontal dashed line) as a result of deviations in total bacterial load across samples.
Figure A 4. Pairwise OTU abundance correlations in the human gut microbiome.

Relationship between total pairwise OTU absolute abundance correlations and (a) temporal and (b) spatial abundance correlations across all pairs of abundant OTUs (mean absolute abundance > 10^{-4}). There is a significant correlation between temporal and total correlations (Pearson’s r = 0.94, p < 1e-10). (c) Total correlations calculated across all pairs of highly abundant OTUs (mean absolute abundance > 10^{-4}) using relative (pink) or absolute (orange) abundances. Boxes show the median and interquartile ranges, with maximum whisker lengths three times the interquartile range. (d) Total pairwise OTU abundance correlations calculated using absolute abundances versus pairwise correlations using relative abundances; each point represents a pair of OTUs. Dashed line indicates the y = x line.
Figure A 5. Correlations of OTU abundances within and between different phyla in the human gut.

(a) Boxplots of total, temporal, spatial and technical correlations of OTU abundances, where pairwise comparisons were made between OTUs within the indicated phyla. Boxes show median and interquartile ranges, with maximum whisker lengths three times the interquartile range. (b-e) Average correlations of OTU abundances within and between different phyla. Colors indicate the average (b) total, (c) temporal, (d) spatial and (e) technical correlations between pairs of OTUs belonging to the indicated phyla.
Figure A 6. Average correlations of OTU abundances within and between different microbial families in the human gut.

Colors indicate the average (a) total, (b) temporal, (c) spatial and (d) technical correlations between pairs of OTUs belonging to the indicated families. The three families belonging to the Bacteroidetes phylum are shown in green.
Figure A 7. Pairwise abundance correlations of gut bacterial OTU abundances within phyla at different phylogenetic distances.

(a) Total, (b) temporal, (c) spatial and (d) technical correlations were calculated for pairs of OTUs belonging to the same phyla, but with different degrees of 16S rRNA sequence dissimilarity. OTU pairs are binned by 16S sequence dissimilarity and mean correlations within each bin are shown for the indicated phyla with error bars denoting the SEM. Pairwise correlations for all abundant OTUs are shown in gray. Proteobacteria were excluded from the analysis due to insufficient sample size.
Figure A 8. Sampling sites from which soil bacteria were collected for 16S rRNA sequencing.

Sites were located around the periphery of a small pond in Central Park, Manhattan.
Figure A 9. Application of DIVERS to a Central Park soil community.

(a) Variance fraction of total bacterial densities attributed to technical noise (N, purple), spatial sampling site (S, blue) and temporal factors (T, red) as calculated by the variance decomposition model. Error bars denote standard deviations based on 1,000 bootstrap iterations. (b) Variance decomposition of individual OTU relative abundances. OTUs are binned by their mean relative abundance across all samples, and stacked bars show the average variance contribution of technical, spatial sampling and temporal sources to OTUs within each bin. Error bars represent the SEM. (c,d) Temporal and spatial variances of soil microbiota abundances calculated using the DIVERS variance decomposition model. Each data point corresponds to the average abundance $(m)$ and (e) temporal or (d) spatial abundance variance $(v)$ of a particular OTU. Variances follow a power law of the form $\propto m^b$ with exponents $b=1.9$ (spatial) and $b=1.6$ (temporal). Only OTUs present in at least two of three samples from each sampling site in at least three sites are shown. (e) Technical noise profiles (variance/mean$^2$) as a function of average abundance across different soil OTUs as calculated using the DIVERS variance decomposition model.
Figure A 10. Decomposition of pairwise OTU abundance correlations in the soil microbiome.

(a) Boxplots of total, temporal, spatial and technical correlations for all pairs of abundant OTUs (average $\log_{10}$ absolute abundance > -3.5). Boxes denote the median and interquartile ranges, with maximum whisker lengths three times the interquartile range. (b) Decomposition of pairwise OTU abundance correlations within and between different phyla. Heatmaps show the average total, temporal, spatial and technical correlations between pairs of OTUs belonging to the indicated phyla.
Table A 1. Metadata for all fecal samples analyzed in this study.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>d16s1r1</td>
<td>1</td>
<td>9:00AM</td>
<td>1</td>
<td>1</td>
<td>41.5</td>
</tr>
<tr>
<td>d16s2r1</td>
<td>1</td>
<td>9:00AM</td>
<td>2</td>
<td>1</td>
<td>85.5</td>
</tr>
<tr>
<td>d16s2r2</td>
<td>1</td>
<td>9:00AM</td>
<td>2</td>
<td>2</td>
<td>61.6</td>
</tr>
<tr>
<td>d17s1r1</td>
<td>2</td>
<td>7:00PM</td>
<td>1</td>
<td>1</td>
<td>49.5</td>
</tr>
<tr>
<td>d17s2r1</td>
<td>2</td>
<td>7:00PM</td>
<td>2</td>
<td>1</td>
<td>56.3</td>
</tr>
<tr>
<td>d17s2r2</td>
<td>2</td>
<td>7:00PM</td>
<td>2</td>
<td>2</td>
<td>52.2</td>
</tr>
<tr>
<td>d18s1r1</td>
<td>3</td>
<td>3:00PM</td>
<td>1</td>
<td>1</td>
<td>46.5</td>
</tr>
<tr>
<td>d18s2r1</td>
<td>3</td>
<td>3:00PM</td>
<td>2</td>
<td>1</td>
<td>63</td>
</tr>
<tr>
<td>d18s2r2</td>
<td>3</td>
<td>3:00PM</td>
<td>2</td>
<td>2</td>
<td>73.9</td>
</tr>
<tr>
<td>d19s1r1</td>
<td>4</td>
<td>7:00PM</td>
<td>1</td>
<td>1</td>
<td>57.4</td>
</tr>
<tr>
<td>d19s2r1</td>
<td>4</td>
<td>7:00PM</td>
<td>2</td>
<td>1</td>
<td>37.1</td>
</tr>
<tr>
<td>d19s2r2</td>
<td>4</td>
<td>7:00PM</td>
<td>2</td>
<td>2</td>
<td>37.5</td>
</tr>
<tr>
<td>d20s1r1</td>
<td>5</td>
<td>12:00PM</td>
<td>1</td>
<td>1</td>
<td>46.4</td>
</tr>
<tr>
<td>d20s2r1</td>
<td>5</td>
<td>12:00PM</td>
<td>2</td>
<td>1</td>
<td>39.6</td>
</tr>
<tr>
<td>d20s2r2</td>
<td>5</td>
<td>12:00PM</td>
<td>2</td>
<td>2</td>
<td>29.3</td>
</tr>
<tr>
<td>d21s1r1</td>
<td>6</td>
<td>10:30PM</td>
<td>1</td>
<td>1</td>
<td>49.9</td>
</tr>
<tr>
<td>d21s2r1</td>
<td>6</td>
<td>10:30PM</td>
<td>2</td>
<td>1</td>
<td>52.4</td>
</tr>
<tr>
<td>d21s2r2</td>
<td>6</td>
<td>10:30PM</td>
<td>2</td>
<td>2</td>
<td>41.5</td>
</tr>
<tr>
<td>d22s1r1</td>
<td>7</td>
<td>11:56PM</td>
<td>1</td>
<td>1</td>
<td>47.7</td>
</tr>
<tr>
<td>d22s2r1</td>
<td>7</td>
<td>11:56PM</td>
<td>2</td>
<td>1</td>
<td>36.7</td>
</tr>
<tr>
<td>d22s2r2</td>
<td>7</td>
<td>11:56PM</td>
<td>2</td>
<td>2</td>
<td>35.8</td>
</tr>
<tr>
<td>d23s1r1</td>
<td>8</td>
<td>4:00PM</td>
<td>1</td>
<td>1</td>
<td>43</td>
</tr>
<tr>
<td>d23s2r1</td>
<td>8</td>
<td>4:00PM</td>
<td>2</td>
<td>1</td>
<td>41</td>
</tr>
<tr>
<td>d23s2r2</td>
<td>8</td>
<td>4:00PM</td>
<td>2</td>
<td>2</td>
<td>44.4</td>
</tr>
<tr>
<td>d24s1r1</td>
<td>9</td>
<td>4:30PM</td>
<td>1</td>
<td>1</td>
<td>64.9</td>
</tr>
<tr>
<td>d24s2r1</td>
<td>9</td>
<td>4:30PM</td>
<td>2</td>
<td>1</td>
<td>38.8</td>
</tr>
<tr>
<td>d24s2r2</td>
<td>9</td>
<td>4:30PM</td>
<td>2</td>
<td>2</td>
<td>35.3</td>
</tr>
<tr>
<td>d25s1r1</td>
<td>10</td>
<td>11:00PM</td>
<td>1</td>
<td>1</td>
<td>59.2</td>
</tr>
<tr>
<td>d25s2r1</td>
<td>10</td>
<td>11:00PM</td>
<td>2</td>
<td>1</td>
<td>39.2</td>
</tr>
<tr>
<td>d25s2r2</td>
<td>10</td>
<td>11:00PM</td>
<td>2</td>
<td>2</td>
<td>56.4</td>
</tr>
<tr>
<td>d26s1r1</td>
<td>11</td>
<td>10:00AM</td>
<td>1</td>
<td>1</td>
<td>36.7</td>
</tr>
<tr>
<td>d26s2r1</td>
<td>11</td>
<td>10:00AM</td>
<td>2</td>
<td>1</td>
<td>58.8</td>
</tr>
<tr>
<td>d26s2r2</td>
<td>11</td>
<td>10:00AM</td>
<td>2</td>
<td>2</td>
<td>68.2</td>
</tr>
<tr>
<td>d28s1r1</td>
<td>13</td>
<td>3:30PM</td>
<td>1</td>
<td>1</td>
<td>57.4</td>
</tr>
<tr>
<td>d28s2r1</td>
<td>13</td>
<td>3:30PM</td>
<td>2</td>
<td>1</td>
<td>56.2</td>
</tr>
<tr>
<td>d28s2r2</td>
<td>13</td>
<td>3:30PM</td>
<td>2</td>
<td>2</td>
<td>45.5</td>
</tr>
<tr>
<td>d29s1r1</td>
<td>14</td>
<td>3:00PM</td>
<td>1</td>
<td>1</td>
<td>58</td>
</tr>
<tr>
<td>d29s2r1</td>
<td>14</td>
<td>3:00PM</td>
<td>2</td>
<td>1</td>
<td>53</td>
</tr>
<tr>
<td>d29s2r2</td>
<td>14</td>
<td>3:00PM</td>
<td>2</td>
<td>2</td>
<td>64.8</td>
</tr>
<tr>
<td>d30s1r1</td>
<td>15</td>
<td>4:00PM</td>
<td>1</td>
<td>1</td>
<td>71.4</td>
</tr>
<tr>
<td>d30s2r1</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>1</td>
<td>16.7</td>
</tr>
<tr>
<td>d30s2r2</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>d31s1r1</td>
<td>16</td>
<td>8:30PM</td>
<td>1</td>
<td>1</td>
<td>32.4</td>
</tr>
<tr>
<td>d31s2r1</td>
<td>16</td>
<td>8:30PM</td>
<td>2</td>
<td>1</td>
<td>63.5</td>
</tr>
<tr>
<td>d31s2r2</td>
<td>16</td>
<td>8:30PM</td>
<td>2</td>
<td>2</td>
<td>50.3</td>
</tr>
<tr>
<td>d32s1r1</td>
<td>17</td>
<td>5:00PM</td>
<td>1</td>
<td>1</td>
<td>62.3</td>
</tr>
<tr>
<td>d32s2r1</td>
<td>17</td>
<td>5:00PM</td>
<td>2</td>
<td>1</td>
<td>46.7</td>
</tr>
<tr>
<td>d32s2r2</td>
<td>17</td>
<td>5:00PM</td>
<td>2</td>
<td>2</td>
<td>68</td>
</tr>
<tr>
<td>d33s1r1</td>
<td>18</td>
<td>12:45PM</td>
<td>1</td>
<td>1</td>
<td>42.6</td>
</tr>
<tr>
<td>d33s2r1</td>
<td>18</td>
<td>12:45PM</td>
<td>2</td>
<td>1</td>
<td>38.1</td>
</tr>
<tr>
<td>d33s2r2</td>
<td>18</td>
<td>12:45PM</td>
<td>2</td>
<td>2</td>
<td>31.8</td>
</tr>
<tr>
<td>d34as1r1</td>
<td>19a</td>
<td>9:00AM</td>
<td>1</td>
<td>1</td>
<td>41.7</td>
</tr>
<tr>
<td>d34as2r1</td>
<td>19a</td>
<td>9:00AM</td>
<td>2</td>
<td>1</td>
<td>46.4</td>
</tr>
<tr>
<td>d34as2r2</td>
<td>19a</td>
<td>9:00AM</td>
<td>2</td>
<td>2</td>
<td>35.2</td>
</tr>
<tr>
<td>d34b1r1</td>
<td>19b</td>
<td>7:00PM</td>
<td>1</td>
<td>1</td>
<td>57.1</td>
</tr>
<tr>
<td>d34b2r1</td>
<td>19b</td>
<td>7:00PM</td>
<td>2</td>
<td>1</td>
<td>46.3</td>
</tr>
<tr>
<td>d34b2r2</td>
<td>19b</td>
<td>7:00PM</td>
<td>2</td>
<td>2</td>
<td>46.2</td>
</tr>
<tr>
<td>d35s1r1</td>
<td>20</td>
<td>4:00PM</td>
<td>1</td>
<td>1</td>
<td>61.9</td>
</tr>
<tr>
<td>d35s2r1</td>
<td>20</td>
<td>4:00PM</td>
<td>2</td>
<td>1</td>
<td>44.6</td>
</tr>
<tr>
<td>d35s2r2</td>
<td>20</td>
<td>4:00PM</td>
<td>2</td>
<td>2</td>
<td>49.0</td>
</tr>
<tr>
<td>d42s1r1</td>
<td>27</td>
<td>7:30PM</td>
<td>1</td>
<td>1</td>
<td>58.3</td>
</tr>
<tr>
<td>d42s2r1</td>
<td>27</td>
<td>7:30PM</td>
<td>2</td>
<td>1</td>
<td>47.7</td>
</tr>
<tr>
<td>d42s2r2</td>
<td>27</td>
<td>7:30PM</td>
<td>2</td>
<td>2</td>
<td>54.9</td>
</tr>
<tr>
<td>d63s1r1</td>
<td>48</td>
<td>10:00PM</td>
<td>1</td>
<td>1</td>
<td>49.5</td>
</tr>
<tr>
<td>d63s2r1</td>
<td>48</td>
<td>10:00PM</td>
<td>2</td>
<td>1</td>
<td>38.2</td>
</tr>
<tr>
<td>d63s2r2</td>
<td>48</td>
<td>10:00PM</td>
<td>2</td>
<td>2</td>
<td>28.4</td>
</tr>
<tr>
<td>d30s3r1</td>
<td>15</td>
<td>4:00PM</td>
<td>3</td>
<td>1</td>
<td>60.8</td>
</tr>
<tr>
<td>d30s4r1</td>
<td>15</td>
<td>4:00PM</td>
<td>4</td>
<td>1</td>
<td>39.5</td>
</tr>
<tr>
<td>d30s5r1</td>
<td>15</td>
<td>4:00PM</td>
<td>5</td>
<td>1</td>
<td>44.7</td>
</tr>
<tr>
<td>d30s6r1</td>
<td>15</td>
<td>4:00PM</td>
<td>6</td>
<td>1</td>
<td>32.1</td>
</tr>
<tr>
<td>d30s7r1</td>
<td>15</td>
<td>4:00PM</td>
<td>7</td>
<td>1</td>
<td>54.1</td>
</tr>
<tr>
<td>d30s8r1</td>
<td>15</td>
<td>4:00PM</td>
<td>8</td>
<td>1</td>
<td>85.1</td>
</tr>
<tr>
<td>d30s9r1</td>
<td>15</td>
<td>4:00PM</td>
<td>9</td>
<td>1</td>
<td>39.7</td>
</tr>
<tr>
<td>d30s10r1</td>
<td>15</td>
<td>4:00PM</td>
<td>10</td>
<td>1</td>
<td>46.7</td>
</tr>
<tr>
<td>d30s11r1</td>
<td>15</td>
<td>4:00PM</td>
<td>11</td>
<td>1</td>
<td>65.0</td>
</tr>
<tr>
<td>d30s12r1</td>
<td>15</td>
<td>4:00PM</td>
<td>12</td>
<td>1</td>
<td>34.7</td>
</tr>
<tr>
<td>d30s13r1</td>
<td>15</td>
<td>4:00PM</td>
<td>13</td>
<td>1</td>
<td>49.0</td>
</tr>
<tr>
<td>d30s14r1</td>
<td>15</td>
<td>4:00PM</td>
<td>14</td>
<td>1</td>
<td>72.7</td>
</tr>
<tr>
<td>d30s2r3</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>3</td>
<td>26.4</td>
</tr>
<tr>
<td>d30s2r4</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>4</td>
<td>21.1</td>
</tr>
<tr>
<td>d30s2r5</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>5</td>
<td>29.1</td>
</tr>
<tr>
<td>d30s2r6</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>6</td>
<td>23.6</td>
</tr>
<tr>
<td>d30s2r7</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>7</td>
<td>15.6</td>
</tr>
<tr>
<td>d30s2r8</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>8</td>
<td>20.9</td>
</tr>
<tr>
<td>d30s2r9</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>9</td>
<td>11.8</td>
</tr>
<tr>
<td>d30s2r10</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>10</td>
<td>21.0</td>
</tr>
<tr>
<td>d30s2r11</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>11</td>
<td>16.4</td>
</tr>
<tr>
<td>d30s2r12</td>
<td>15</td>
<td>4:00PM</td>
<td>2</td>
<td>12</td>
<td>11.2</td>
</tr>
</tbody>
</table>
Table A.2. 16S sequencing primers utilized in the study.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16Sf_501</td>
<td>AATGATACGCGCCGACCACCAGATCTACAC TAGATCGC TATGGTAATT GT GTGYCAGCMGCCGCGGTAA</td>
</tr>
<tr>
<td>16Sf_502</td>
<td>AATGATACGCGCCGACCACCAGATCTACAC CTCTCTAT TATGGTAATT GT GTGYCAGCMGCCGCGGTAA</td>
</tr>
<tr>
<td>16Sf_503</td>
<td>AATGATACGCGCCGACCACCAGATCTACAC TATCCTCT TATGGTAATT GT GTGYCAGCMGCCGCGGTAA</td>
</tr>
<tr>
<td>16Sf_504</td>
<td>AATGATACGCGCCGACCACCAGATCTACAC AGAGTAGA TATGGTAATT GT GTGYCAGCMGCCGCGGTAA</td>
</tr>
<tr>
<td>16Sf_505</td>
<td>AATGATACGCGCCGACCACCAGATCTACAC GTAAGGAG TATGGTAATT GT GTGYCAGCMGCCGCGGTAA</td>
</tr>
<tr>
<td>16Sf_506</td>
<td>AATGATACGCGCCGACCACCAGATCTACAC ACTGCATA TATGGTAATT GT GTGYCAGCMGCCGCGGTAA</td>
</tr>
<tr>
<td>16Sf_507</td>
<td>AATGATACGCGCCGACCACCAGATCTACAC AAGGAGTA TATGGTAATT GT GTGYCAGCMGCCGCGGTAA</td>
</tr>
<tr>
<td>16Sf_508</td>
<td>AATGATACGCGCCGACCACCAGATCTACAC CTAAGCCT TATGGTAATT GT GTGYCAGCMGCCGCGGTAA</td>
</tr>
<tr>
<td>16Sr_701</td>
<td>CAAGCAGAAGACCGCATACGAGAT TCAGCCTTA AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_702</td>
<td>CAAGCAGAAGACCGCATACGAGAT CTAGTACG AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_703</td>
<td>CAAGCAGAAGACCGCATACGAGAT TTCTGCCT AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_704</td>
<td>CAAGCAGAAGACCGCATACGAGAT GCTCAGGA AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_705</td>
<td>CAAGCAGAAGACCGCATACGAGAT AGGAGTCC AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_706</td>
<td>CAAGCAGAAGACCGCATACGAGAT CATGCCTA AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_707</td>
<td>CAAGCAGAAGACCGCATACGAGAT GTAGAGAG AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_708</td>
<td>CAAGCAGAAGACCGCATACGAGAT CCTCTCTG AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_709</td>
<td>CAAGCAGAAGACCGCATACGAGAT AGCGTAGC AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_710</td>
<td>CAAGCAGAAGACCGCATACGAGAT CAGCCTCG AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_711</td>
<td>CAAGCAGAAGACCGCATACGAGAT TGCCCTCT AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16Sr_712</td>
<td>CAAGCAGAAGACCGCATACGAGAT CCTCTCCT AGTCAGTCAG CC GGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16S_read1</td>
<td>TATGGTAATTGTGTYCAGCMGCCGCGGTAA</td>
</tr>
<tr>
<td>16S_read2</td>
<td>AGTCAGTCGCGCGACTACNVGGGTWTCTAAT</td>
</tr>
<tr>
<td>16S_index1</td>
<td>ATTAGAWACCCBNCTAGTCCGCTGACTGACT</td>
</tr>
</tbody>
</table>
Table A. OTUs with > 60% spatial or > 80% temporal abundance variance relative to the total variance.

<table>
<thead>
<tr>
<th>OTU</th>
<th>Phylum</th>
<th>Class</th>
<th>Order</th>
<th>Family</th>
<th>Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Firmicutes (98%)</td>
<td>Clostridia (98%)</td>
<td>Clostridiales (98%)</td>
<td>Ruminococcaceae (94%)</td>
<td>Clostridium IV (68%)</td>
</tr>
<tr>
<td>33</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Ruminococcus (100%)</td>
</tr>
<tr>
<td>48</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Peptostreptococcaceae (100%)</td>
<td>Intestiobacter (88%)</td>
</tr>
<tr>
<td>71</td>
<td>Verrucomicrobia (100%)</td>
<td>Verrucomicrobiales (100%)</td>
<td>Verrucomicrobiaceae (100%)</td>
<td>Akkermansia (100%)</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>Firmicutes (100%)</td>
<td>Erysipelotrichia (100%)</td>
<td>Erysipelotrichales (100%)</td>
<td>Erysipelotrichaceae (100%)</td>
<td>Turicibacter (100%)</td>
</tr>
<tr>
<td>101</td>
<td>Proteobacteria (100%)</td>
<td>Gammaproteobacteria (100%)</td>
<td>Pasteurellales (100%)</td>
<td>Pasteurellaceae (100%)</td>
<td>Haemophilus (69%)</td>
</tr>
<tr>
<td>122</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Peptostreptococcaceae (100%)</td>
<td>Terrisporobacter (100%)</td>
</tr>
<tr>
<td>6</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Faecalibacterium (93%)</td>
</tr>
<tr>
<td>11</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Fusocatenibacter (96%)</td>
</tr>
<tr>
<td>12</td>
<td>Actinobacteria (100%)</td>
<td>Actinobacteria (100%)</td>
<td>Bifidobacteriales (100%)</td>
<td>Bifidobacteriaceae (100%)</td>
<td>Bifidobacterium (100%)</td>
</tr>
<tr>
<td>14</td>
<td>Firmicutes (100%)</td>
<td>Erysipelotrichia (100%)</td>
<td>Erysipelotrichiales (100%)</td>
<td>Erysipelotrichaceae (100%)</td>
<td>Clostridium XVIII (75%)</td>
</tr>
<tr>
<td>22</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Ruminococcus (100%)</td>
</tr>
<tr>
<td>24</td>
<td>Bacteroides (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidales (100%)</td>
<td>Bacteroidiaceae (100%)</td>
<td>Bacteroides (100%)</td>
</tr>
<tr>
<td>25</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Lachnospiraceae_incertae_sedis (75%)</td>
</tr>
<tr>
<td>27</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Peptostreptococcaceae (100%)</td>
<td>Romboutsia (96%)</td>
</tr>
<tr>
<td>28</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Lachnospiraceae_incertae_sedis (92%)</td>
</tr>
<tr>
<td>34</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Lachnospiraceae_incertae_sedis (68%)</td>
</tr>
<tr>
<td>35</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Faecalibacterium (77%)</td>
</tr>
<tr>
<td>36</td>
<td>Proteobacteria (100%)</td>
<td>Gammaproteobacteria (100%)</td>
<td>Enterobacteriales (100%)</td>
<td>Enterobacteriaceae (100%)</td>
<td>Escherichia/Shigella (79%)</td>
</tr>
<tr>
<td>37</td>
<td>Firmicutes (100%)</td>
<td>Bacillia (100%)</td>
<td>Lactobacillales (100%)</td>
<td>Enterococaceae (100%)</td>
<td>Enterococcus (100%)</td>
</tr>
<tr>
<td>39</td>
<td>Proteobacteria (100%)</td>
<td>Gammaproteobacteria (100%)</td>
<td>Enterobacteriales (100%)</td>
<td>Enterobacteriaceae (100%)</td>
<td>Salmonella (98%)</td>
</tr>
<tr>
<td>41</td>
<td>Bacteroides (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidales (100%)</td>
<td>Bacteroidiaceae (100%)</td>
<td>Bacteroides (100%)</td>
</tr>
<tr>
<td>43</td>
<td>Firmicutes (100%)</td>
<td>Negativicutes (100%)</td>
<td>Selenomonadaceae (100%)</td>
<td>Acidaminococcaceae (100%)</td>
<td>Phascolarctobacterium (100%)</td>
</tr>
<tr>
<td>44</td>
<td>Bacteroides (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidales (100%)</td>
<td>Bacteroidaceae (100%)</td>
<td>Bacteroides (100%)</td>
</tr>
<tr>
<td>46</td>
<td>Firmicutes (99%)</td>
<td>Clostridia (98%)</td>
<td>Clostridiales (98%)</td>
<td>Ruminococcaceae (74%)</td>
<td>Flavonifractor (19%)</td>
</tr>
<tr>
<td>50</td>
<td>Bacteroides (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidales (100%)</td>
<td>Rikenellaceae (100%)</td>
<td>Alistipes (100%)</td>
</tr>
<tr>
<td>52</td>
<td>Firmicutes (96%)</td>
<td>Clostridia (95%)</td>
<td>Clostridiales (95%)</td>
<td>Ruminococcaceae (78%)</td>
<td>Anaerotruncus (39%)</td>
</tr>
<tr>
<td>54</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (95%)</td>
<td>Clostridium XIVa (33%)</td>
</tr>
<tr>
<td>55</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Roseburia (54%)</td>
</tr>
<tr>
<td>56</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Oscillibacter (70%)</td>
</tr>
<tr>
<td>57</td>
<td>Firmicutes (100%)</td>
<td>Bacillia (100%)</td>
<td>Lactobacillales (100%)</td>
<td>Streptococcaceae (100%)</td>
<td>Lactococcus (100%)</td>
</tr>
<tr>
<td>59</td>
<td>Firmicutes (100%)</td>
<td>Negativicutes (100%)</td>
<td>Selenomonadaceae (100%)</td>
<td>Acidaminococcaceae (100%)</td>
<td>Acidaminococcus (100%)</td>
</tr>
<tr>
<td>60</td>
<td>Bacteroides (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidales (100%)</td>
<td>Bacteroidaceae (100%)</td>
<td>Bacteroides (100%)</td>
</tr>
<tr>
<td>61</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (98%)</td>
<td>Coprococcus (25%)</td>
</tr>
<tr>
<td>62</td>
<td>Bacteroides (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidales (100%)</td>
<td>Bacteroidaceae (100%)</td>
<td>Bacteroides (100%)</td>
</tr>
<tr>
<td>64</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Clostridium XIVa (44%)</td>
</tr>
<tr>
<td>67</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (99%)</td>
<td>Anaeroplasma (21%)</td>
</tr>
<tr>
<td>68</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Clostridium IV (65%)</td>
</tr>
<tr>
<td>69</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Lachnospiraceae_incertae_sedis (47%)</td>
</tr>
<tr>
<td>70</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Fusocatenibacter (58%)</td>
</tr>
<tr>
<td>73</td>
<td>Firmicutes (99%)</td>
<td>Clostridia (98%)</td>
<td>Clostridiales (98%)</td>
<td>Clostridiales_incertae_Sedis XIII (63%)</td>
<td>Mogibacterium (45%)</td>
</tr>
<tr>
<td>76</td>
<td>Firmicutes (100%)</td>
<td>Erysipelotrichia (100%)</td>
<td>Erysipelotrichiales (100%)</td>
<td>Erysipelotrichaceae (100%)</td>
<td>Clostridium XVIII (100%)</td>
</tr>
<tr>
<td>79</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (99%)</td>
<td>Lachnобacterium (49%)</td>
</tr>
<tr>
<td>81</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Clostridium IV (100%)</td>
</tr>
<tr>
<td>85</td>
<td>Firmicutes (100%)</td>
<td>Erysipelotrichia (100%)</td>
<td>Erysipelotrichales (100%)</td>
<td>Erysipelotrichaceae (100%)</td>
<td>Coprobacillus (84%)</td>
</tr>
<tr>
<td>88</td>
<td>Bacteroides (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidales (100%)</td>
<td>Porphyromonadaceae (100%)</td>
<td>Parabacteroides (100%)</td>
</tr>
<tr>
<td>89</td>
<td>Firmicutes (98%)</td>
<td>Clostridia (98%)</td>
<td>Clostridiales (98%)</td>
<td>Ruminococcaceae (93%)</td>
<td>Clostridium IV (81%)</td>
</tr>
<tr>
<td>90</td>
<td>Proteobacteria (100%)</td>
<td>Burkholderiales (100%)</td>
<td>Burkholderiaceae (100%)</td>
<td>Sutterellaceae (100%)</td>
<td>Parasutterella (100%)</td>
</tr>
<tr>
<td></td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (99%)</td>
<td>Flavonifractor (77%)</td>
</tr>
<tr>
<td>---</td>
<td>------------------</td>
<td>------------------</td>
<td>----------------------</td>
<td>-----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>94</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (99%)</td>
<td>Flavonifractor (77%)</td>
</tr>
<tr>
<td>97</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (98%)</td>
<td>Dorea (46%)</td>
</tr>
<tr>
<td>99</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (85%)</td>
<td>Sporobacter (41%)</td>
</tr>
<tr>
<td>103</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Clostridium XIVa (57%)</td>
</tr>
<tr>
<td>106</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (99%)</td>
<td>Oscillibacter (98%)</td>
</tr>
<tr>
<td>108</td>
<td>Bacteroidetes (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidiales (100%)</td>
<td>Rikenellaceae (100%)</td>
<td>Alistipes (100%)</td>
</tr>
<tr>
<td>109</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Flavonifractor (100%)</td>
</tr>
<tr>
<td>112</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (93%)</td>
<td>Butyrivibrio (24%)</td>
</tr>
<tr>
<td>114</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Flavonifractor (67%)</td>
</tr>
<tr>
<td>116</td>
<td>Firmicutes (100%)</td>
<td>Negativicutes (100%)</td>
<td>Selenomonadaceae (100%)</td>
<td>Veillonellaceae (100%)</td>
<td>Veillonella (100%)</td>
</tr>
<tr>
<td>120</td>
<td>Firmicutes (95%)</td>
<td>Clostridia (93%)</td>
<td>Clostridiales (93%)</td>
<td>Lachnospiraceae (50%)</td>
<td>Clostridium XIVb (50%)</td>
</tr>
<tr>
<td>130</td>
<td>Bacteroidetes (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidales (100%)</td>
<td>Bacteroidaceae (100%)</td>
<td>Bacteroides (100%)</td>
</tr>
<tr>
<td>136</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (99%)</td>
<td>Anaerobacter (25%)</td>
</tr>
<tr>
<td>140</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Sporobacter (78%)</td>
</tr>
<tr>
<td>305</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Clostridium XIVa (32%)</td>
</tr>
<tr>
<td>354</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Ruminococcaceae (100%)</td>
<td>Faecalibacterium (100%)</td>
</tr>
<tr>
<td>374</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Ruminococcus2 (100%)</td>
</tr>
<tr>
<td>392</td>
<td>Actinobacteria (100%)</td>
<td>Actinobacteria (100%)</td>
<td>Bifidobacteriales (100%)</td>
<td>Bifidobacteriaceae (100%)</td>
<td>Bifidobacterium (95%)</td>
</tr>
</tbody>
</table>
### Table A 4. Metadata for all soil samples analyzed in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Site No.</th>
<th>Rep No.</th>
<th>Date</th>
<th>Time</th>
<th>Sample Weight (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil1</td>
<td>The Pool in Central Park, Manhattan</td>
<td>1</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>261.5</td>
</tr>
<tr>
<td>Soil2</td>
<td>The Pool in Central Park, Manhattan</td>
<td>2</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>258.1</td>
</tr>
<tr>
<td>Soil3</td>
<td>The Pool in Central Park, Manhattan</td>
<td>2</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>234.1</td>
</tr>
<tr>
<td>Soil4</td>
<td>The Pool in Central Park, Manhattan</td>
<td>3</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>266.1</td>
</tr>
<tr>
<td>Soil5</td>
<td>The Pool in Central Park, Manhattan</td>
<td>4</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>278.9</td>
</tr>
<tr>
<td>Soil6</td>
<td>The Pool in Central Park, Manhattan</td>
<td>4</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>272.4</td>
</tr>
<tr>
<td>Soil7</td>
<td>The Pool in Central Park, Manhattan</td>
<td>5</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>252</td>
</tr>
<tr>
<td>Soil8</td>
<td>The Pool in Central Park, Manhattan</td>
<td>5</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>263.1</td>
</tr>
<tr>
<td>Soil9</td>
<td>The Pool in Central Park, Manhattan</td>
<td>6</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>266.8</td>
</tr>
<tr>
<td>Soil10</td>
<td>The Pool in Central Park, Manhattan</td>
<td>6</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>236.2</td>
</tr>
<tr>
<td>Soil11</td>
<td>The Pool in Central Park, Manhattan</td>
<td>7</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>256</td>
</tr>
<tr>
<td>Soil12</td>
<td>The Pool in Central Park, Manhattan</td>
<td>8</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>266.1</td>
</tr>
<tr>
<td>Soil13</td>
<td>The Pool in Central Park, Manhattan</td>
<td>8</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>244</td>
</tr>
<tr>
<td>Soil14</td>
<td>The Pool in Central Park, Manhattan</td>
<td>9</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>651.3</td>
</tr>
<tr>
<td>Soil15</td>
<td>The Pool in Central Park, Manhattan</td>
<td>10</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>318.7</td>
</tr>
<tr>
<td>Soil16</td>
<td>The Pool in Central Park, Manhattan</td>
<td>10</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>268.7</td>
</tr>
<tr>
<td>Soil17</td>
<td>The Pool in Central Park, Manhattan</td>
<td>10</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>277.5</td>
</tr>
<tr>
<td>Soil18</td>
<td>The Pool in Central Park, Manhattan</td>
<td>11</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>299.9</td>
</tr>
<tr>
<td>Soil19</td>
<td>The Pool in Central Park, Manhattan</td>
<td>12</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>282.6</td>
</tr>
<tr>
<td>Soil20</td>
<td>The Pool in Central Park, Manhattan</td>
<td>12</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>260.4</td>
</tr>
<tr>
<td>Soil21</td>
<td>The Pool in Central Park, Manhattan</td>
<td>13</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>292.3</td>
</tr>
<tr>
<td>Soil22</td>
<td>The Pool in Central Park, Manhattan</td>
<td>14</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>360.7</td>
</tr>
<tr>
<td>Soil23</td>
<td>The Pool in Central Park, Manhattan</td>
<td>14</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>278</td>
</tr>
<tr>
<td>Soil24</td>
<td>The Pool in Central Park, Manhattan</td>
<td>15</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>341.5</td>
</tr>
<tr>
<td>Soil25</td>
<td>The Pool in Central Park, Manhattan</td>
<td>15</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>319.8</td>
</tr>
<tr>
<td>Soil26</td>
<td>The Pool in Central Park, Manhattan</td>
<td>16</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>261</td>
</tr>
<tr>
<td>Soil27</td>
<td>The Pool in Central Park, Manhattan</td>
<td>16</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>267.8</td>
</tr>
<tr>
<td>Soil28</td>
<td>The Pool in Central Park, Manhattan</td>
<td>17</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>250</td>
</tr>
<tr>
<td>Soil29</td>
<td>The Pool in Central Park, Manhattan</td>
<td>18</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>281.8</td>
</tr>
<tr>
<td>Soil30</td>
<td>The Pool in Central Park, Manhattan</td>
<td>18</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>253.3</td>
</tr>
<tr>
<td>Soil31</td>
<td>The Pool in Central Park, Manhattan</td>
<td>19</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>267.1</td>
</tr>
<tr>
<td>Soil32</td>
<td>The Pool in Central Park, Manhattan</td>
<td>20</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>329.2</td>
</tr>
<tr>
<td>Soil33</td>
<td>The Pool in Central Park, Manhattan</td>
<td>20</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>323.4</td>
</tr>
<tr>
<td>Soil34</td>
<td>The Pool in Central Park, Manhattan</td>
<td>20</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>429.3</td>
</tr>
<tr>
<td>Soil35</td>
<td>The Pool in Central Park, Manhattan</td>
<td>21</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>376.9</td>
</tr>
<tr>
<td>Soil36</td>
<td>The Pool in Central Park, Manhattan</td>
<td>22</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>259.1</td>
</tr>
<tr>
<td>Soil37</td>
<td>The Pool in Central Park, Manhattan</td>
<td>22</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>273.5</td>
</tr>
<tr>
<td>Soil38</td>
<td>The Pool in Central Park, Manhattan</td>
<td>23</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>289.1</td>
</tr>
<tr>
<td>Soil39</td>
<td>The Pool in Central Park, Manhattan</td>
<td>24</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>329</td>
</tr>
<tr>
<td>Soil40</td>
<td>The Pool in Central Park, Manhattan</td>
<td>24</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>367.2</td>
</tr>
<tr>
<td>Soil41</td>
<td>The Pool in Central Park, Manhattan</td>
<td>25</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>370.6</td>
</tr>
<tr>
<td>Soil42</td>
<td>The Pool in Central Park, Manhattan</td>
<td>26</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>272.4</td>
</tr>
<tr>
<td>Soil43</td>
<td>The Pool in Central Park, Manhattan</td>
<td>26</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>294.6</td>
</tr>
<tr>
<td>Soil44</td>
<td>The Pool in Central Park, Manhattan</td>
<td>27</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>347.4</td>
</tr>
<tr>
<td>Soil45</td>
<td>The Pool in Central Park, Manhattan</td>
<td>28</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>262.7</td>
</tr>
<tr>
<td>Soil46</td>
<td>The Pool in Central Park, Manhattan</td>
<td>28</td>
<td>2</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>251.5</td>
</tr>
<tr>
<td>Soil47</td>
<td>The Pool in Central Park, Manhattan</td>
<td>28</td>
<td>1</td>
<td>15-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>276.9</td>
</tr>
<tr>
<td>Soil48</td>
<td>The Pool in Central Park, Manhattan</td>
<td>1</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>327.2</td>
</tr>
<tr>
<td>Soil49</td>
<td>The Pool in Central Park, Manhattan</td>
<td>1</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>307.3</td>
</tr>
<tr>
<td>Soil50</td>
<td>The Pool in Central Park, Manhattan</td>
<td>2</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>258.9</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>3</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>263</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------</td>
<td>----</td>
<td>----</td>
<td>-----------</td>
<td>----------------</td>
<td>-----</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>3</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>238.9</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>4</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>263.7</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>5</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>292.1</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>5</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>308</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>5</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>293.1</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>6</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>266.6</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>7</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>296.1</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>7</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>274.4</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>8</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>282.2</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>9</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>371.1</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>9</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>347.2</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>10</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>272</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>10</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>380.1</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>11</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>400.6</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>11</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>407.8</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>12</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>326.9</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>13</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>476.4</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>13</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>392.1</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>14</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>328.4</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>15</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>315.5</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>15</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>276.6</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>15</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>313.8</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>16</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>261.4</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>17</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>226.8</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>17</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>103.9</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>18</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>274.1</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>19</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>251.8</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>19</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>250.2</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>20</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>294.8</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>20</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>314.4</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>21</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>342.6</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>21</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>338.8</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>22</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>249.4</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>23</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>294.5</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>23</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>274.4</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>24</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>292.6</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>25</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>359.7</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>25</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>330.8</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>25</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>279</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>26</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>249.1</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>27</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>293.1</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>27</td>
<td>2</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>290.5</td>
</tr>
<tr>
<td>Soil</td>
<td>The Pool in Central Park, Manhattan</td>
<td>28</td>
<td>1</td>
<td>22-Jun-18</td>
<td>9:30AM-11:45AM</td>
<td>271.2</td>
</tr>
</tbody>
</table>
APPENDIX B: SUPPLEMENTARY FIGURES AND TABLES FOR CHAPTER 4

Author statement

Some figures and tables have been adapted from the article:
Ji, BW, Sheth, RU, Dixit, Purushottam D, Vitkup, D. Macroecological relationships in the

Authors

Brian W. Ji$^{1,3}$, Ravi U. Sheth$^{1,3}$, Purushottam D. Dixit$^1$ & Dennis Vitkup$^{1,2*}$

Author affiliations

$^1$Department of Systems Biology, Columbia University, New York, NY, USA. $^2$Department of
Biomedical Informatics, Columbia University, New York, NY, USA. $^3$These authors contributed
equally.
Daily growth rates of gut microbiota were defined as the logarithm of successive abundance ratios, \( \mu = \log \left( \frac{X(t+1)}{X(t)} \right) \), where \( X(t) \) is relative abundance of an OTU on day \( t \). Growth rates averaged over the community are consistent with a Laplace distribution of the form: 

\[
p(\mu) = \frac{1}{2b} \exp(-|\mu|/b)
\]

(a) humans (\( b = 0.83 \pm 0.1, 0.67 \pm 0.1, 0.71 \pm 0.07, 0.73 \pm 0.05 \), A, B, M3, F4 respectively; mean \( \pm \) s.d. Methods) and (b) mice (\( b = 0.82 \pm 0.1, 0.67 \pm 0.03 \), LFPP and HFHS diets respectively mean \( \pm \) s.d. across individual mice). (c) Daily growth rates of gut microbiota defined at different taxonomic resolutions. Growth rates were calculated as the logarithm of successive abundance ratios, \( \mu = \log \left( \frac{X(t+1)}{X(t)} \right) \), where \( X(t) \) corresponds to the sum of all OTU abundances falling within the same assigned taxonomy on day \( t \). OTUs were defined at the level of 97% sequence similarity in 16S rRNA. (d) Distribution of standardized daily growth rates in the human gut microbiome. Daily growth rates were first normalized by their standard deviations within respective individual OTUs. The distribution of resulting standardized growth rates are shown for humans A, B, M3 and F4. For all panels, solid lines indicate MLE fits to the data.
Figure B 2. Growth rate distributions in diverse ecological communities and economic systems.

(a) Annual growth rate distributions of North American bird populations\textsuperscript{9}, marine species abundances\textsuperscript{11}, publicly-traded company sales\textsuperscript{13} and university R&D expenditures\textsuperscript{22}. Figures were adapted from their original text. Distributions of company sales and R&D expenditures were replotted for companies with initial dollar sales of $4^{15}$ and universities with large R&D expenditures (see refs. 13 and 22 for details). Lines are provided for visual purposes only. (b) Residence time distributions of species belonging to diverse ecosystems. Figures were adapted from ref. 5, with original data comprising North American breeding bird species, estuarine fish species, and plant species collected from both prairie and forest ecosystems. Lines are provided for visual purposes only. (c) Long-term behavior of bird species abundances\textsuperscript{9} and North Atlantic fish stock abundances\textsuperscript{23}. Figures were adapted from their original text.
Figure B 3. Long-term dynamics of gut microbiota abundances in humans and mice.

(a,b) Mean-squared displacements of log OTU abundances $\langle \delta^2(\Delta t) \rangle$ as a function of time $\Delta t$. Dashed lines are fits to the equation $\langle \delta^2(\Delta t) \rangle \propto \Delta t^{2H}$, where $H$ is the Hurst exponent characterizing the diffusion process. Hurst exponents are $H = 0.07 \pm 0.03$, $0.10 \pm 0.04$, $0.08 \pm 0.02$, $0.1 \pm 0.07$ for humans A, B, M3, and F4 respectively, and $H = 0.08 \pm 0.02$, $0.19 \pm 0.02$ for LFPP and HFHS mice (mean ± s.d., Methods).
Figure B 4. Long-term behavior of individual OTU abundances.

(a) Mean-squared-displacements of individual OTU abundances $\langle \delta^2(\Delta t) \rangle_k$ as a function of time $\Delta t$. Dashed lines are fits to the equation $\langle \delta^2(\Delta t) \rangle_k \propto \Delta t^{2H}$. Panels correspond to OTUs analyzed from human A. (b) Distributions of Hurst exponents for individual OTUs in humans. The Hurst exponent for the entire community is indicated by a dashed line.
Figure B 5. Residence and return times of gut microbiota in humans and mice.

Distributions of residence and return times for (a) human and (b) mouse gut microbiota. Solid lines are fits to the data using a power law with exponential tail of the form $p(t) \propto t^{-\alpha} e^{-\lambda t}$. In humans, power law exponents are $\alpha_{res} = 2.3 \pm 0.04, 2.2 \pm 0.05, 2.2 \pm 0.07, 2.14 \pm 0.08$ for residence times and $\alpha_{ret} = 1.1 \pm 0.02, 1.15 \pm 0.03, 1.2 \pm 0.05, 1.09 \pm 0.07, 1$ for return times (A, B, M3, F4 respectively; mean ± s.d., Methods). In mice, $\alpha_{res} = 2.2 \pm 0.04, 2.2 \pm 0.03$ and $\alpha_{ret} = 0.72 \pm 0.03, 0.67 \pm 0.06$ for the LFPP and HFHS groups respectively (mean ± s.d. across individual mice).
Figure B 6. Taylor’s power law in the human and mouse gut microbiome.

Temporal variances $\sigma_X^2$ as a function of average species abundances $\langle X \rangle$ in (a) human and (b) mouse gut microbiomes. Dashed lines are least-squares fits to Taylor’s power law of the form $\sigma_X^2 \propto \langle X \rangle^\beta$. Each dot represents the mean and temporal variance of a single OTU. (a) Power law exponents in human individuals are $\beta = 1.66 \pm 0.09, 1.60 \pm 0.08, 1.71 \pm 0.07, 1.71 \pm 0.07$ for A, B, M3, and F4 respectively (mean ± s.d., Methods). Colored dots denote OTUs described in the main text whose abundances on any day exceeded the average abundance over all other days by over 25-fold (Table D.1). (b) Power law exponents in mice are $\beta = 1.49 \pm 0.02$ and $1.86 \pm 0.07$ for LFPP and HFHS diets respectively (mean ± s.d. across individual mice). (c) Relative abundances of spiking OTUs identified in (a) for two human subjects (A and B) whose lifestyles were closely documented over the time series. Major events affecting the gut microbiota of both individuals may be identified, including the travel of individual A to a developing country near day 100 and an enteric infection in individual B near day 150.
Table B 1. OTUs in humans with a single day abundance exceeding the average abundance over all others days by over 25-fold.

<table>
<thead>
<tr>
<th>Human</th>
<th>OTU</th>
<th>Phylum</th>
<th>Class</th>
<th>Order</th>
<th>Family</th>
<th>Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>13</td>
<td>Proteobacteria</td>
<td>Gammaproteobacteria (97%)</td>
<td>Enterobacteriales (85%)</td>
<td>Enterobacteriaceae (85%)</td>
<td>Escherichia/Shigella (37%)</td>
</tr>
<tr>
<td>A</td>
<td>25</td>
<td>Firmicutes (98%)</td>
<td>Clostridia (98%)</td>
<td>Clostridales (98%)</td>
<td>Peptostreptococcaceae (93%)</td>
<td>Clostridium XI (71%)</td>
</tr>
<tr>
<td>B</td>
<td>13</td>
<td>Proteobacteria</td>
<td>Gammaproteobacteria (100%)</td>
<td>Enterobacteriales (85%)</td>
<td>Enterobacteriaceae (85%)</td>
<td>Escherichia/Shigella (37%)</td>
</tr>
<tr>
<td>B</td>
<td>59</td>
<td>Actinobacteria (100%)</td>
<td>Actinobacteria (100%)</td>
<td>Coriobacteriales (100%)</td>
<td>Coriobacteriaceae (100%)</td>
<td>Eggerthella (100%)</td>
</tr>
<tr>
<td>B</td>
<td>95</td>
<td>Firmicutes (71%)</td>
<td>Clostridia (59%)</td>
<td>Clostridiales (55%)</td>
<td>Ruminococcaceae (23%)</td>
<td>Gemmiger (12%)</td>
</tr>
<tr>
<td>B</td>
<td>110</td>
<td>Firmicutes (67%)</td>
<td>Clostridia (60%)</td>
<td>Clostridiales (53%)</td>
<td>Ruminococcaceae (20%)</td>
<td>Butyrivibrio (11%)</td>
</tr>
<tr>
<td>B</td>
<td>5928</td>
<td>Firmicutes (99%)</td>
<td>Clostridia (99%)</td>
<td>Clostridiales (99%)</td>
<td>Lachnospiraceae (99%)</td>
<td>Oribacterium (60%)</td>
</tr>
<tr>
<td>M3</td>
<td>10</td>
<td>Verrucomicrobia (100%)</td>
<td>Verrucomicrobiae (100%)</td>
<td>Verrucomicrobiales (100%)</td>
<td>Verrucomicrobiaceae (100%)</td>
<td>Akkermansia (100%)</td>
</tr>
<tr>
<td>M3</td>
<td>11</td>
<td>Bacteroidetes (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidales (100%)</td>
<td>Bacteroidaceae (100%)</td>
<td>Bacteroides (100%)</td>
</tr>
<tr>
<td>M3</td>
<td>12</td>
<td>Proteobacteria (100%)</td>
<td>Gammaproteobacteria (98%)</td>
<td>Enterobacteriales (85%)</td>
<td>Enterobacteriaceae (85%)</td>
<td>Cronobacter (30%)</td>
</tr>
<tr>
<td>M3</td>
<td>31</td>
<td>Bacteroidetes (100%)</td>
<td>Bacteroidia (100%)</td>
<td>Bacteroidales (100%)</td>
<td>Porphyromonadaceae (100%)</td>
<td>Porphyromonas (99%)</td>
</tr>
<tr>
<td>M3</td>
<td>41</td>
<td>Firmicutes (53%)</td>
<td>Clostridia (45%)</td>
<td>Clostridiales (41%)</td>
<td>Clostridiales_Incertae_Sedis XIII (7%)</td>
<td>Anaerovorax (7%)</td>
</tr>
<tr>
<td>M3</td>
<td>44</td>
<td>Firmicutes (86%)</td>
<td>Clostridia (86%)</td>
<td>Clostridiales (86%)</td>
<td>Ruminococcaceae (85%)</td>
<td>Oscillospira (77%)</td>
</tr>
<tr>
<td>M3</td>
<td>71</td>
<td>Firmicutes (100%)</td>
<td>Bacilli (100%)</td>
<td>Lactobacillales (100%)</td>
<td>Streptococcaceae (100%)</td>
<td>Streptococcus (100%)</td>
</tr>
<tr>
<td>M3</td>
<td>75</td>
<td>Firmicutes (94%)</td>
<td>Clostridia (94%)</td>
<td>Clostridiales (94%)</td>
<td>Ruminococcaceae (93%)</td>
<td>Subdoligranulum (78%)</td>
</tr>
<tr>
<td>F4</td>
<td>14</td>
<td>Bacteroidetes (100%)</td>
<td>Bacteroidia (99%)</td>
<td>Bacteroidales (99%)</td>
<td>Prevotellaceae (99%)</td>
<td>Prevotella (99%)</td>
</tr>
<tr>
<td>F4</td>
<td>2643</td>
<td>Firmicutes (100%)</td>
<td>Clostridia (100%)</td>
<td>Clostridiales (100%)</td>
<td>Lachnospiraceae (100%)</td>
<td>Ruminococcus2 (94%)</td>
</tr>
</tbody>
</table>
Table B.2. OTUs in LFPP and HFHS mice with significantly higher or lower variability than expected from Taylor’s power law.

<table>
<thead>
<tr>
<th>Mouse</th>
<th>OTU</th>
<th>Direction</th>
<th>Phylum</th>
<th>Class</th>
<th>Order</th>
<th>Family</th>
<th>Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>HFHS1</td>
<td>4</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridia</td>
<td>Clostridiales</td>
<td>Ruminococcaceae</td>
<td>Butyricoccus</td>
</tr>
<tr>
<td>HFHS1</td>
<td>33</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridia</td>
<td>Clostridiales</td>
<td>Ruminococcaceae</td>
<td>Butyricoccus</td>
</tr>
<tr>
<td>HFHS1</td>
<td>36</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Erysipelotrichales</td>
<td>Erysipelotrichaceae</td>
<td>Allobaculum</td>
</tr>
<tr>
<td>HFHS1</td>
<td>55</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Ruminococcaceae</td>
<td>Lachnospiraceae</td>
<td>Robinsoniella</td>
</tr>
<tr>
<td>HFHS2</td>
<td>67</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Erysipelotrichales</td>
<td>Erysipelotrichaceae</td>
<td>Allobaculum</td>
</tr>
<tr>
<td>HFHS2</td>
<td>264</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Ruminococcaceae</td>
<td>Lachnospiraceae</td>
<td>Robinsoniella</td>
</tr>
<tr>
<td>HFHS2</td>
<td>183</td>
<td>Low</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Erysipelotrichales</td>
<td>Erysipelotrichaceae</td>
<td>Allobaculum</td>
</tr>
<tr>
<td>HFHS3</td>
<td>5</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Ruminococcaceae</td>
<td>Lachnospiraceae</td>
<td>Clostridium XIVa</td>
</tr>
<tr>
<td>HFHS3</td>
<td>21</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Ruminococcaceae</td>
<td>Lachnospiraceae</td>
<td>Ruminococcus2</td>
</tr>
<tr>
<td>HFHS3</td>
<td>33</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Ruminococcaceae</td>
<td>Butlerivibrio</td>
<td>Pseudoflavonifactor</td>
</tr>
<tr>
<td>HFHS3</td>
<td>243</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Ruminococcaceae</td>
<td>Lachnospiraceae</td>
<td>Clostridium IV</td>
</tr>
<tr>
<td>LFPP1</td>
<td>237</td>
<td>Low</td>
<td>Bacteroides</td>
<td>Bacteroides</td>
<td>Bacteroidales</td>
<td>Porphyromonadaceae</td>
<td>Barnesiella</td>
</tr>
<tr>
<td>LFPP1</td>
<td>364</td>
<td>Low</td>
<td>Bacteroides</td>
<td>Bacteroides</td>
<td>Bacteroidales</td>
<td>Porphyromonadaceae</td>
<td>Barnesiella</td>
</tr>
<tr>
<td>LFPP2</td>
<td>93</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Clostridiales</td>
<td>Clostridaceae 1</td>
<td>Anaerophorobacter</td>
</tr>
<tr>
<td>LFPP2</td>
<td>99</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Clostridiales</td>
<td>Clostridaceae 1</td>
<td>Anaerophorobacter</td>
</tr>
<tr>
<td>LFPP2</td>
<td>276</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Clostridiales</td>
<td>Clostridaceae 1</td>
<td>Anaerophorobacter</td>
</tr>
<tr>
<td>LFPP2</td>
<td>237</td>
<td>Low</td>
<td>Bacteroides</td>
<td>Bacteroides</td>
<td>Bacteroidales</td>
<td>Porphyromonadaceae</td>
<td>Barnesiella</td>
</tr>
<tr>
<td>LFPP3</td>
<td>16</td>
<td>High</td>
<td>Firmicutes</td>
<td>Clostridiales</td>
<td>Clostridiales</td>
<td>Clostridaceae 1</td>
<td>Anaerophorobacter</td>
</tr>
<tr>
<td>LFPP3</td>
<td>22</td>
<td>Low</td>
<td>Bacteroides</td>
<td>Bacteroides</td>
<td>Bacteroidales</td>
<td>Porphyromonadaceae</td>
<td>Barnesiella</td>
</tr>
<tr>
<td>LFPP3</td>
<td>118</td>
<td>Low</td>
<td>Bacteroides</td>
<td>Bacteroides</td>
<td>Bacteroidales</td>
<td>Clostridium 4</td>
<td>Barnesiella</td>
</tr>
<tr>
<td>LFPP3</td>
<td>237</td>
<td>Low</td>
<td>Bacteroides</td>
<td>Bacteroides</td>
<td>Bacteroidales</td>
<td>Porphyromonadaceae</td>
<td>Barnesiella</td>
</tr>
<tr>
<td>LFPP3</td>
<td>364</td>
<td>Low</td>
<td>Bacteroides</td>
<td>Bacteroides</td>
<td>Bacteroidales</td>
<td>Porphyromonadaceae</td>
<td>Barnesiella</td>
</tr>
</tbody>
</table>
APPENDIX C: FULL DESCRIPTION OF VARIANCE AND COVARIANCE DECOMPOSITION MODELS FOR CHAPTER 3

Author statement

Some passages and figures have been adapted or quoted verbatim from the article:

Authors

Brian W. Ji$^{1,2\dagger}$, Ravi U. Sheth$^{1,2\dagger}$, Purushottam D. Dixit$^1$, Andrew Kaufman$^1$, Harris H. Wang$^{1,3*}$ & Dennis Vitkup$^{1,4*}$

Author Affiliations

$^1$Department of Systems Biology, Columbia University, New York, NY, USA.
$^2$Integrated Program in Cellular, Molecular, and Biomedical Studies, Columbia University, New York, NY, USA.
$^3$Department of Pathology and Cell Biology, Columbia University, New York, NY, USA
$^4$Department of Biomedical Informatics, Columbia University, New York, NY, USA.
$\dagger$These authors contributed equally to this work.
Variance decomposition model

Overview

Let $X_t$ be a random variable denoting the abundance of a given bacterial operational taxonomic unit (OTU) $i$ measured from 16S rRNA sequencing. Although we focus on single OTU abundances, the following also applies for total bacterial abundances in each sample. We let the measured abundances of OTU $i$ collected at different time points of a time series study reflect draws from an underlying distribution $p(X_t)$, which we refer to as the marginal distribution of $X_t$. We assume that there are three contributions to the total observed variability of $X_t$. The first corresponds to any set of temporal factors that systematically change from one day to another to drive changes in OTU abundances over time. This may encompass environmental factors, interspecies interactions and competition, and neutral drift in abundances. The second contribution reflects heterogeneity in the abundance of bacteria across different spatial locations in a given environment. Notably, this spatial sampling variability is inherent to some time series studies such as those of the gut microbiome, where measurements from fecal samples collected at different time points necessarily come from different spatial locations. We take the spatial sampling variability to reflect variation arising from differences in niche size and availability or random dispersal of taxa abundances. The third corresponds to experimental noise associated with bacterial DNA extraction from samples, PCR amplification and sequencing itself. We collectively refer to these experimental sources of variability as technical noise. Our goal is to derive expressions for each of these sources of variability and demonstrate how one may estimate them from experiments.

Decomposing the variance in microbiota abundances
We first demonstrate how we can use the law of total variance to decompose measured bacterial abundance variances into the three contributions described in the previous section.

**Defining the space and time variables**

As bacterial abundances change over time, we would expect that abundances of the same OTU $i$ measured at the same time point across different spatial locations in the environment would be more similar to each other than those collected from different time points. However, even at the same time point, measured abundances will not be identical due to both variation across different spatial locations and technical noise. Mathematically, the variability in abundances of OTU $i$ at a fixed time $T = t$ defines a conditional random variable $X_i^t$ with distribution $p(X_i^t | T = t)$, where $T$ is a time-associated random variable that captures the collective state of all time-associated factors and has underlying distribution $p(T)$. This conditional distribution itself may change when $T$ realizes different values at different time points. Importantly, however, when conditioned on a particular time point $T = t$, the variance of the conditional distribution $\text{Var}(X_i^t)$ reflects only spatial heterogeneity and technical noise.
At a given point in time, we may also choose a location from which to collect a sample to sequence. We can therefore define another random variable $X_{i}^{t,s}$ with probability distribution $p(X_{i}|T = t, S = s)$ representing the abundance of OTU $i$ measured from this fixed time point $T = t$ and spatial location $S = s$. Here, $S$ is a space-associated random variable with distribution $p(S|T = t)$ that at a given point in time, changes with the particular spatial location from which OTU $i$ is collected and measured. Conditioning on both space and time, we have eliminated any biological sources of variability and the variance of OTU $i$, $Var(X_{i}^{t,s})$ simply reflects technical noise. The distributions $p(X_{i})$, $p(X_{i}|T = t)$, $p(X_{i}|T = t, S = s)$ and their hierarchical relationships are illustrated in Fig. S1, with the fecal microbiome shown as the model ecosystem.

Variance decomposition
Using the law of total variance, we now decompose the total abundance variance of \( X_i \) into components associated with time, spatial sampling location and technical noise. In the following, \( E \) and \( Var \) denote the expectation and variance of a random variable respectively, and subscripts denote the underlying distribution \((p(T) \text{ or } p(S|T))\) with respect to which the operation is performed. Beginning with the definition of the variance of \( X_i \),

\[
Var(X_i) = E(X_i^2) - [E(X_i)]^2
\]

\[
= E_T E_{S|T} E(X_i^2|S,T) - [E_T E_{S|T} E(X_i|S,T)]^2
\]

\[
= E_T E_{S|T} Var(X_i|S,T) + E_T E_{S|T} [E(X_i|S,T)]^2 - [E_T E_{S|T} E(X_i|S,T)]^2
\]

\[
= E_T E_{S|T} Var(X_i|S,T) + E_T Var_{S|T} E(X_i|S,T) + E_T [E_{S|T} E(X_i|S,T)]^2 - [E_T E_{S|T} E(X_i|S,T)]^2
\]

Thus,

\[
Var(X_i) = E_T E_{S|T} Var(X_i|S,T) + E_T Var_{S|T} E(X_i|S,T) + Var_T E_{S|T} E(X_i|S,T)
\]

The terms in the last line correspond to the contributions of technical, spatial sampling, and temporal factors to the total variance of \( X_i \), which we denote with the symbols \( \langle \sigma^2_N \rangle_{S,T}, \langle \sigma^2_S \rangle_T \) and \( \sigma^2_T \) respectively. Equation (2) is simply the law of total variance generalized to multiple conditional random variables. Indeed, the right-most term can be recognized as the variance of \( X_i \) explained by the time random variable \( T \). The second term reflects the spatial sampling variance of OTU abundances conditioned on time, then averaged over time. The first term is simply the technical variability conditioned on a spatial location and time point, and then jointly averaged over space and time.

Model-driven experimental approach
We now describe the approach we use to estimate the three contributions to total abundance variability. We show mathematically how our model can be used to estimate each of these terms with only a minimal set of experiments. Notably, the approach we adopt is a generalization of the dual reporter method originally described by Elowitz et al. (Swain et al., 2002) used to separate intrinsic versus extrinsic sources of noise in the gene expression profiles of single cells (Fu and Pachter, 2016; Hilfinger and Paulsson, 2011; Swain et al., 2002).

**Experimental setup**

Extending the notation described above, we denote $X_i$, $Y_i$ and $Z_i$ to be random variables representing abundances of an OTU $i$ made from three separate measurements at each point in time from the community. Let us assume that the abundances $X_i$ and $Y_i$ are made from the same exact location (by sequencing the sample twice), whereas $Z_i$ is measured from an independent location. As $X_i$, $Y_i$ and $Z_i$ correspond to the same bacterial OTU, their marginal distributions are equivalent. Importantly, however, because the abundances $X_i$, $Y_i$ and $Z_i$ are sampled together at the same time points, there exists a covariance structure driven by shared but unknown temporal factors tending to cause their abundances to collectively increase or decrease from one day to the next. Therefore, although $X_i$, $Y_i$ and $Z_i$ are identically distributed, they are not independent. In addition, by letting $X_i$ and $Y_i$ correspond to abundances measured not only at the same time point but also from the same spatial location, the covariance between $X_i$ and $Y_i$ is also driven by shared spatial factors that result in similar OTU abundances across different locations.
We can break the described covariance structures by conditioning abundances on time and spatial location. Specifically, when conditioning abundances on time \( T = t \), the pairs \( X_i^t, Z_i^t \) and \( Y_i^t, Z_i^t \) have become sets of independent draws from the distribution \( p(X_i|T = t) \). Experimentally, independence is achieved if \( Z_i^t \) is sampled from a location in the community independent from \( X_i^t \) and \( Y_i^t \). Note that while \( X_i^t \) and \( Y_i^t \) have the same underlying distribution \( p(X_i^t|T = t) \), they are not independent as their values covary across space. However, further conditioning of \( X_i^t \) and \( Y_i^t \) on spatial location results in the conditional random variables \( X_i^{ts} \) and \( Y_i^{ts} \) which are indeed independent draws from the distribution \( p(X_i|T = t, S = s) \). The assumption of independence is reasonable, as \( X_i^{ts} \) and \( Y_i^{ts} \) are simply technical replicates.

Therefore, our replicate sampling protocol goes as follows: at each time point, we make three abundance measurements for all bacterial OTUs \( 1 \ldots N \). Two of these abundance measurements \( (X_i \text{ and } Y_i) \) are made from the same spatial location in a given environment. The third \( (Z_i) \) is measured from a separate, independent location. From an experimental standpoint, \( X_i \) and \( Y_i \) correspond to technical replicates while \( X_i, Z_i \) and \( Y_i, Z_i \) correspond to spatial replicates. Here, \( t \) in \( 1 \ldots n \) with \( n \) being the total number of time points for which bacterial samples are collected.

**Derivation of statistical estimators for variance decomposition**

We can now derive the statistical estimators for each of the terms in equation (2) using the complete hierarchical model described previously. We begin by showing that under the specified model, the first two moments of the marginal distributions of \( X_i, Y_i, \) and \( Z_i \) are indeed identical.
We have now laid the groundwork for the following derivations of statistical estimators for each of the terms in equation (2). This is the primary result of the variance decomposition model.

**Variance:** By the law of total variance,

\[
\text{Var}(X_i) = E_T \text{Var}(X_i | S, T) + E_T \text{Var}(X_i | S, T) + \text{Var}_T E(S | T) E(X_i | S, T)
\]

\[
= E_T \text{Var}(Y_i | S, T) + E_T \text{Var}(Z_i | T) + \text{Var}_T E(Y_i | S, T) = \text{Var}(Y_i)
\]

\[
\text{Var}(X_i) = E_T \text{Var}(X_i | T) + \text{Var}_T E(X_i | T) = E_T \text{Var}(Z_i | T) + \text{Var}_T E(Z_i | T) = \text{Var}(Z_i)
\]

We have now laid the groundwork for the following derivations of statistical estimators for each of the terms in equation (2). This is the primary result of the variance decomposition model.

**Variance associated with time**

\[
\sigma_T^2 = \text{Var}_T E(S | T) E(X_i | S, T) = \text{Var}_T E(X_i | T)
\]

\[
= E_T [E(X_i | T)]^2 - [E_T E(X_i | T)]^2
\]

\[
= E_T [E(X_i | T)E(Z_i | T)] - E_T E(X_i | T)E_T E(Z_i | T)
\]

\[
= E_T E(X_i Z_i | T) - E_T E(X_i | T)E_T E(Z_i | T)
\]

\[
= E(X_i Z_i) - E(X_i) E(Z_i) = \text{Cov}(X_i, Z_i)
\]

Hence, the time-associated variability is simply the covariance between \(X_i\) and \(Z_i\), which can be easily estimated as:

\[
\hat{\sigma}_T^2 = \frac{1}{n-1} \sum_{t=1}^{n} (x_i^t - \bar{x}_i)(z_i^t - \bar{z}_i)
\]
Variance associated with spatial sampling location

\[ \langle \sigma_{S}^{2} \rangle_T = E_T \text{Var}_{S|T} E(X_i|S,T) = E_T E_{S|T}[E(X_i|S,T)]^2 - E_T[E_{S|T}E(X_i|S,T)]^2 \]

\[ = E_T E_{S|T}[E(X_i|S,T)]^2 - E_T[E(X_i|T)]^2 \]

\[ = E_T E_{S|T}[E(X_i|S,T)E(Y_i|S,T)] - E_T[E(Z_i|T)][E(Y_i|T)] \]

\[ = E_T E_{S|T}E(X_iY_i|S,T) - E_T E(Z_iY_i|T) \]

\[ = E(X_iY_i) - E(Z_iY_i) \]

\[ = E(X_iY_i - Z_iY_i) - E(X_i)E(Y_i) + E(Z_i)E(Y_i) \]

\[ = E((X_i - Z_i)Y_i) - E(X_i - Z_i)E(Y_i) = Cov(X_i - Z_i, Y_i) \]  \hfill (9)

Similar to time, the spatial sampling-associated variance reduces to the covariance between \(X_i-Z_i\) and \(Y_i\) and is estimated by:

\[ \langle \sigma_{S}^{2} \rangle_T = \frac{1}{n - 1} \sum_{i=1}^{n} [(x_i^T - z_i^T) - (\bar{x}_i - \bar{z}_i)](y_i^T - \bar{y}_i) \]  \hfill (10)

Technical noise
Finally, the technical variability is simply the variance of the difference between $X_i$ and $Y_i$, which is estimated as:

$$\langle \sigma^2_{N_{S,T}} \rangle_{S,T} = \frac{1}{2(2n-1)} \sum_{t=1}^{n} [(x_i^t - y_i^t) - (\bar{x}_i - \bar{y}_i)]^2$$  

(12)

**Generalizing the hierarchy**

In the sections above, we have implicitly imposed a hierarchy in our model. Namely, quantities in each term of equation (2) are first averaged with respect to $p(X_i|S,T)$ followed by $p(S|T)$ and finally $p(T)$. One can also imagine reversing this hierarchy; that is, averaging with respect to $p(T|S)$ second, followed by $p(S)$ last. Experimentally, this would correspond to drawing temporal replicates from a fixed spatial location, then averaging quantities over various locations. In ecosystems such as the human gut microbiome, the hierarchy described in earlier sections arises naturally, as $p(S|T)$ and $p(T)$ are experimentally accessible from fecal samples, while $p(T|S)$ and $p(S)$ are not. In other words, it is only possible to draw spatial replicates from a fixed time point,
then average bacterial abundance measurements over multiple time points. It is not possible to measure abundances from the same spatial location on fecal samples obtained from two different days. However, in other microbial communities, such as those found in the soil, one can imagine an alternative hierarchical sampling protocol in which different spatial locations in the ecosystem are each sampled on two separate days. Equation (2) may then be used to decompose bacterial abundance variances using this inverted hierarchy:

\[
\text{Var}(X_i) = E_\text{S}E_{T|S}\text{Var}(X_i|S,T) + E_\text{S}\text{Var}_{T|S}\text{E}(X_i|S,T) + \text{Var}_{S}\text{E}(X_i|S,T)\\
\text{Technical (}(\sigma^2_N)_{S,T}\text{)} \quad \text{Temporal (}(\sigma^2_T)_{S}\text{)} \quad \text{Spatial sampling (}(\sigma^2_S)\text{)}
\]

The first term reflecting technical noise remains the same. The second term now reflects a temporal variance that is averaged over sampling locations. The third reflects the variance explained by spatial sampling location. Indeed, we used this inverted hierarchy to carry out sampling of a soil microbial community in Central Park. We also note that the spatial sampling variability here may be defined rather loosely. For example, if one were to collect two human microbiome samples at different time points from a large cohort of individuals and sequence one of these two samples twice, the spatial sampling variability could be referred to as inter-individual variability while the temporal variability would correspond to intra-individual variability, with the noise term remaining the same. Finally, we remark that abundance measurements may be made from multiple different spatial locations at multiple different time points, with multiple sequencing replicates performed at each time and location. While experimentally much more demanding, one may arrive at the different variance contributions directly by following the prescription of equation (2) and using the typical estimators for mean and variance. However, our hierarchical replicate sampling protocol makes such data collection unnecessary.
Covariance decomposition model

Overview

We now extend our variance decomposition model for single OTUs to a generalized covariance decomposition for all pairs of OTUs. We let $X_i$ and $X_j$ denote the abundances of OTUs $i$ and $j$ measured together from the same sample (i.e. sequenced from the same spatial location in a given environment). As with the case of single OTU variances, we assume that the total abundance covariance between OTUs $i$ and $j$ (across different samples collected over time) may be attributed to underlying temporal, spatial and technical sources. Intuitively, the temporal contribution results from the covariation in overall abundances (averaged over all spatial locations in the community) of OTUs $i$ and $j$ from one day to the next. In addition, however, species abundances may be correlated across different spatial locations at a given time point, which is captured by the spatial contribution to the total covariance. Finally, technical factors may result in correlated noise, potentially arising from sources such as similar DNA extraction efficiencies or primer and amplification biases. Again, our goal is to derive expressions for each of these covariance sources and demonstrate how one may estimate them experimentally using the same protocol described above.
Figure C 2. Decomposing the covariance in pairs of bacterial species abundances.

Total and conditional joint distributions

The total abundance covariance of OTUs $i$ and $j$ may be calculated from the simple experiment: draw a single sample from a random spatial location in the environment at each time point and sequence the abundances of $i$ and $j$. Mathematically, we may consider the bivariate random variable $\tilde{X}$ with components comprising the measured abundances $p(X_i, X_j)$ and define a total joint distribution $p(X_{i}, X_{j})$. We refer to this as the total joint distribution because in the data collection process, we have marginalized over time, space and technical noise. In contrast, one can imagine a second experiment where at a given time point, multiple samples are obtained across various spatial locations in the community. This defines another distribution, the conditional joint distribution $p(X_i, X_j|T = t)$ for some fixed time point $t$, where variances and covariances now reflect both underlying spatial factors as well as technical noise. Finally, by fixing both time and
sampling location $S = s$ and re-sequencing the extracted DNA multiple times, a third distribution $p(X_i, X_j | T = t, S = s)$ can be defined. Here, variances and covariances reflect purely technical sources. The hierarchical relationships of these distributions are illustrated in Fig. S2. We show in the next section how the total covariance $Cov(X_i, X_j)$ between OTUs $i$ and $j$ may be decomposed by making use of the described conditional joint distributions.

Covariance decomposition

The total covariance corresponding to the distribution $p(X_i, X_j)$ can be written as:

$$Cov(X_i, X_j) = E(X_i X_j) - E(X_i) E(X_j)$$

$$= E_T E_{S|T} E(X_i X_j | S, T) - E(X_i) E(X_j)$$

$$= E_T E_{S|T} Cov(X_i, X_j | S, T) + E_T E_{S|T} [E(X_i | S, T) E(X_j | S, T)] - E(X_i) E(X_j)$$

$$= E_T E_{S|T} Cov(X_i, X_j | S, T) + E_T Cov_{S|T}(E(X_i | S, T), E(X_j | S, T))$$

$$+ E_T [E_{S|T} E(X_i | S, T) E_{S|T} E(X_j | S, T)] - E(X_i) E(X_j)$$

(14)

Mirroring the variance decomposition, we arrive at:

$$Cov(X_i, X_j) = E_T E_{S|T} Cov(X_i, X_j | S, T) + E_T Cov_{S|T}(E(X_i | S, T), E(X_j | S, T)) + Cov_T(E(X_i | T), E(X_j | T))$$

(15)

Note that we can obtain correlations by simply dividing each term in equation (15) by marginal standard deviations.

Derivation of statistical estimators for covariance decomposition

Let $X_i Y_i$, and $Z_i$ and $X_j Y_j$, and $Z_j$ denote abundances of OTUs $i$ and $j$ respectively as described in Section 1.3.1. That is, the pairs $(X_i, Z_i)$ and $(Y_i, Z_i)$ correspond to spatial replicates while $(X_i, Y_i)$
denote technical replicates. As before, when conditioning on time \( T = t \), the bivariate random variable pairs \( \tilde{X}^t = (X_i^t, X_j^t) \) and \( \tilde{Z}^t = (Z_i^t, Z_j^t) \), and \( \tilde{Y}^t = (Y_i^t, Y_j^t) \) and \( \tilde{Z}^t = (Z_i^t, Z_j^t) \) correspond to two i.i.d. draws from the same underlying distribution \( p(X_i, X_j|T = t) \), with a covariance matrix structured by both spatial and technical sources. We will now define two additional and equivalent conditional joint distributions \( p(X_i, Z_j|T = t) \) and \( p(Y_i, Z_j|T = t) \) where the abundance of OTU \( i \) from one spatial replicate \( X_i^t \) or \( Y_i^t \) is paired with the abundance of OTU \( j \) from the second spatial replicate \( Z_i^t \). Note here that while the conditional marginal distributions of \( p(X_i, Z_j|T = t) \) and \( p(Y, Z_j|T = t) \) are identical to those of the distribution \( p(X_i, X_j|T = t) \) as demonstrated in the previous section, the random variables \( X_i^t \) and \( Z_j^t \), and \( Y_i^t \) and \( Z_j^t \) are independent of one another when conditioned on time, as abundances from each pair come from different spatial locations and sequencing realizations. Along similar lines, the bivariate random variables \( \tilde{X}^{t,s} = (X_i^{t,s}, X_j^{t,s}) \) and \( \tilde{Y}^{t,s} = (Y_i^{t,s}, Y_j^{t,s}) \) correspond to random variables drawn from the distribution \( p(X_i, X_j|T = t, S = s) \), where correlations between \( i \) and \( j \) are driven purely by technical sources. Again, we may preserve conditional marginal distributions while eliminating covariances by defining the distribution \( p(X_i, Y_j|T = t, S = s) \). With this in mind, we will now derive statistical estimators for each of the terms in equation (15).

Covariance associated with time
Analogous to the variance decomposition, the time-associated covariance can be easily estimated as:

\[\sigma_i^* \sigma_j^* = \text{Cov}_T(E(X_i|T), E(X_j|T))\]

\[= E_T[E(X_i|T)E(X_j|T)] - E_T[E(X_i|T)E(X_j|T)]\]

\[= E_T[E(X_i|T)E(Z_j|T)] - E_T[E(X_i|T)E(Z_j|T)]\]

\[= E_T[E(X_iZ_j|T) - E(X_i)E(Z_j)]\]

\[= \text{Cov}(X_i, Z_j)\] (16)

Covariance associated with spatial sampling location

\[\langle \sigma_i^* \sigma_j^* \rangle_T = E_T \text{Cov}_S(E(X_i|S, T), E(X_j|S, T))\]

\[= E_T E_S[T][E(X_i|S, T)E(X_j|S, T)] - E_T E_S[T][E(X_i|S, T)E(X_j|S, T)]\]

\[= E_T E_S[T][E(X_i|S, T)E(X_j|S, T)] - E_T E(X_i|T)E(X_j|T)]\]

\[= E_T E_S[T][E(X_i|S, T)E(Y_j|S, T)] - E_T E(X_i|T)E(Y_j|T)]\]

\[= E_T E_S[T]E(X_iY_j|S, T) - E_T E(Z,Y_j|T)\]

\[= E(X_iY_j) - E(Z,Y_j)\]

\[= E(X_iY_j) - E(Z,Y_j) - E(X_i)E(Y_j) + E(Z_i)E(Y_j)\]

\[= E((X_i - Z_i)Y_j) + E(X_i - Z_i)E(Y_j)\]

\[= \text{Cov}(X_i - Z_i, Y_j)\] (18)

The space-associated covariance is thus given by:

\[\langle \sigma^i \sigma^j \rangle_S = \frac{1}{n-1} \sum_{t=1}^{n} [(x_i^t - \bar{x}_i) - (\bar{x}_t - \bar{y}_i)](y_j^t - \bar{y}_j)\] (19)
Covariance associated with technical noise

\[
\langle \sigma_i^2 \sigma_j^2 \rangle_{S,T} = E_T E_{S|T} \text{Cov}(X_i, X_j|S, T)
\]

\[
= E_T E_{S|T} E(X_i X_j|S, T) - E_T E_{S|T} [E(X_i|S, T)E(X_j|S, T)]
\]

\[
= \frac{1}{2} [E_T E_{S|T} E(X_i X_j|S, T) - E_T E_{S|T} [E(X_i|S, T)E(Y_j|S, T)]
\]

\[
- E_T E_{S|T} [E(Y_i|S, T)E(X_j|S, T)] + E_T E_{S|T} E(Y_i Y_j|S, T)]
\]

\[
= \frac{1}{2} [E_T E_{S|T} E(X_i X_j|S, T) - E_T E_{S|T} E(X_i Y_j|S, T)
\]

\[
- E_T E_{S|T} E(Y_i X_j|S, T) + E_T E_{S|T} E(Y_i Y_j|S, T)]
\]

\[
= \frac{1}{2} [E(X_i X_j) - E(X_i Y_j) - E(Y_i X_j) + E(Y_i Y_j)]
\]

\[
= \frac{1}{2} [E(X_i X_j - X_i Y_j + Y_i X_j + Y_i Y_j) - E(X_i)E(X_j) + E(X_i)E(Y_j) - E(Y_i)E(X_j) + E(Y_i)E(Y_j)]
\]

\[
= \frac{1}{2} [E(X_i - Y_i)(X_j - Y_j) - E(X_i - Y_i)E(X_j - Y_j)]
\]

\[
= \frac{1}{2} \text{Cov}(X_i - Y_i, X_j - Y_j)
\]

Finally, the covariance associated with technical sources is estimated as:

\[
\langle \sigma_i^2 \sigma_j^2 \rangle_{S,T} = \frac{1}{2(n-1)} \sum_{i=1}^{n} [(x_i^r - y_i^r) - (\bar{x}_i - \bar{y}_i)][(x_j^r - y_j^r) - (\bar{x}_j - \bar{y}_j)]
\]

Covariances induced by conversion to absolute abundances

Finally, we demonstrate that conversion from relative to absolute abundances typically results in higher covariances for OTU pairs measured in absolute abundances. These higher covariances stem from the variance of total bacterial densities measured from sample to sample that may induce correlations in OTU pairs whose relative abundances are otherwise uncorrelated or even negatively correlated. Let \( R_i \) and \( R_j \) denote the relative abundances of OTUs \( i \) and \( j \) measured from a single
sample. Let us denote $A$ to be the total bacterial abundance density (in units of DNA copies per mg of environmental sample matter). Note that the absolute abundances $X_i$ and $X_j$ are simply calculated as $AR_i$ and $AR_j$ respectively. We will assume that the relative abundances $R_i$ and $R_j$ are independent of $A$. Then, we may write:

$$
\text{Cov}(AR_i, AR_j) = E(A^2 R_i R_j) - E(A)^2 E(R_i) E(R_j)
$$

$$
= E(A^2)E(R_i R_j) - E(A)^2 E(R_i) E(R_j)
$$

$$
= E(A^2)[\text{Cov}(R_i, R_j) + E(R_i)E(R_j)] - E(A)^2 E(R_i) E(R_j)
$$

$$
= E(A^2)\text{Cov}(R_i, R_j) + E(R_i)E(R_j)[E(A^2) - E(A)^2]
$$

$$
= E(A^2)\text{Cov}(R_i, R_j) + E(R_i)E(R_j)\text{Var}(A)
$$

(22)

References


Author statement

Some passages and figures have been adapted or quoted verbatim from the article: Lipids are a major growth limitation in cells deficient for electron acceptors, 2018. In Preparation.

Authors

Brian W. Ji¹, Zhaoqi Li², Purushottam D. Dixit¹, Lucas B. Sullivan², Matthew G. Vander Heiden²* & Dennis Vitkup¹,³*

Author affiliations

¹Department of Systems Biology, Columbia University, New York, NY, USA. ²The Koch Institute for Integrative Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. ³Department of Biomedical Informatics, Columbia University, New York, NY, USA.
Figure D 1. Lipid species concentrations in serum-delipidated media.

(a,b) Relative concentrations of palmitate and oleate in lipid-replete (red) and lipid-depleted media conditions (green, purple).
Figure D 2. Lipid synthesis rates in lipid-depleted media conditions.

(a) Relative ion counts of intracellular palmitate in HeLa cells as a function of time in lipid-replete and lipid-depleted conditions. (b) HeLa palmitate synthesis rates in lipid-replete versus lipid-depleted conditions. (c) Proliferation rates of HeLa cells grown in the presence or absence of extracellular lipids treated with the fatty acid synthase inhibitor GSK2194069.
Figure D 3. Oxidative lipid synthesis in lipid-replete or lipid-depleted conditions.

(a) Oxygen consumption rates in HeLa cells grown in the presence or absence of extracellular lipids. (b) Intracellular NAD+/NADH ratio in HeLa cells grown in the presence or absence of extracellular lipids. (c) Intracellular citrate pools in HeLa cells. (d) M+2 citrate ion counts derived from U-C^{13} as a function of time in the indicated conditions.
Table D 1. List of biomass metabolites in the genome-scale model of human metabolism.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Metabolite classification</th>
<th>Stoichiometry in biomass production reaction (mmol/g-DW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>'adp[c]'</td>
<td>Nucleotide</td>
<td>29.0585</td>
</tr>
<tr>
<td>'ala-L[c]'</td>
<td>Amino acid</td>
<td>-0.5056</td>
</tr>
<tr>
<td>'arg-L[c]'</td>
<td>Amino acid</td>
<td>-0.3593</td>
</tr>
<tr>
<td>'asn-L[c]'</td>
<td>Amino acid</td>
<td>-0.2794</td>
</tr>
<tr>
<td>'asp-L[c]'</td>
<td>Amino acid</td>
<td>-0.3526</td>
</tr>
<tr>
<td>'atp[c]'</td>
<td>Nucleotide</td>
<td>-29.1123</td>
</tr>
<tr>
<td>'chsterol[c]'</td>
<td>Lipid/sterol</td>
<td>-0.0204</td>
</tr>
<tr>
<td>'clpn_hs[c]'</td>
<td>Lipid/sterol</td>
<td>-0.0117</td>
</tr>
<tr>
<td>'ctp[c]'</td>
<td>Nucleotide</td>
<td>-0.039</td>
</tr>
<tr>
<td>'cys-L[c]'</td>
<td>Amino acid</td>
<td>-0.0466</td>
</tr>
<tr>
<td>'datp[n]'</td>
<td>Nucleotide</td>
<td>-0.0132</td>
</tr>
<tr>
<td>'dctp[n]'</td>
<td>Nucleotide</td>
<td>-0.0094</td>
</tr>
<tr>
<td>'dgtp[n]'</td>
<td>Nucleotide</td>
<td>-0.0099</td>
</tr>
<tr>
<td>'dttp[n]'</td>
<td>Nucleotide</td>
<td>-0.0131</td>
</tr>
<tr>
<td>'g6p[c]'</td>
<td>Carbohydrate</td>
<td>-0.2752</td>
</tr>
<tr>
<td>'gln-L[c]'</td>
<td>Amino acid</td>
<td>-0.326</td>
</tr>
<tr>
<td>'glu-L[c]'</td>
<td>Amino acid</td>
<td>-0.3859</td>
</tr>
<tr>
<td>'gly[c]'</td>
<td>Amino acid</td>
<td>-0.5389</td>
</tr>
<tr>
<td>'gtp[c]'</td>
<td>Nucleotide</td>
<td>-0.0361</td>
</tr>
<tr>
<td>'h2o[c]'</td>
<td>Inorganic</td>
<td>-29.0585</td>
</tr>
<tr>
<td>'h[c]'</td>
<td>Inorganic</td>
<td>29.0585</td>
</tr>
<tr>
<td>'his-L[c]'</td>
<td>Amino acid</td>
<td>-0.1264</td>
</tr>
<tr>
<td>'ile-L[c]'</td>
<td>Amino acid</td>
<td>-0.2861</td>
</tr>
<tr>
<td>'leu-L[c]'</td>
<td>Amino acid</td>
<td>-0.5455</td>
</tr>
<tr>
<td>'lys-L[c]'</td>
<td>Amino acid</td>
<td>-0.5921</td>
</tr>
<tr>
<td>'met-L[c]'</td>
<td>Amino acid</td>
<td>-0.153</td>
</tr>
<tr>
<td>'pail_hs[c]'</td>
<td>Lipid/sterol</td>
<td>-0.0233</td>
</tr>
<tr>
<td>'pchol_hs[c]'</td>
<td>Lipid/sterol</td>
<td>-0.1545</td>
</tr>
<tr>
<td>'phol_hs[c]'</td>
<td>Lipid/sterol</td>
<td>-0.0554</td>
</tr>
<tr>
<td>'pglyc_hs[c]'</td>
<td>Lipid/sterol</td>
<td>-0.0029</td>
</tr>
<tr>
<td>'phe-L[c]'</td>
<td>Amino acid</td>
<td>-0.2595</td>
</tr>
<tr>
<td>'pi[c]'</td>
<td>Inorganic</td>
<td>29.0585</td>
</tr>
<tr>
<td>'pro-L[c]'</td>
<td>Amino acid</td>
<td>-0.4125</td>
</tr>
<tr>
<td>'ps_hs[c]'</td>
<td>Lipid/sterol</td>
<td>-0.0058</td>
</tr>
<tr>
<td>'ser-L[c]'</td>
<td>Amino acid</td>
<td>-0.3925</td>
</tr>
<tr>
<td>Substance</td>
<td>Category</td>
<td>Value</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------</td>
<td>---------</td>
</tr>
<tr>
<td>'sphmyln_hs[c]'</td>
<td>Lipid/sterol</td>
<td>-0.0175</td>
</tr>
<tr>
<td>'thr-L[c]'</td>
<td>Amino acid</td>
<td>-0.3127</td>
</tr>
<tr>
<td>'try-L[c]'</td>
<td>Amino acid</td>
<td>-0.133</td>
</tr>
<tr>
<td>'tyr-L[c]'</td>
<td>Amino acid</td>
<td>-0.1597</td>
</tr>
<tr>
<td>'utp[c]'</td>
<td>Nucleotide</td>
<td>-0.0534</td>
</tr>
<tr>
<td>'val-L[c]'</td>
<td>Amino acid</td>
<td>-0.3526</td>
</tr>
</tbody>
</table>