# THE HUMAN MICROBIOME AND INTERACTIONS WITH DISEASE STATES, MEDICAL INTERVENTIONS, AND THE METABOLOME.

by

LUKE KENNETH URSELL

B.Sc., University of Florida, 2010

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Biochemistry

2014

This thesis entitled:

The Human Microbiome and Interactions with Disease States, Medical Interventions, and the Metabolome
written by Luke Kenneth Ursell
has been approved for the Department of Biochemistry

_____

Rob Knight

_____

William Old

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Ursell, Luke Kenneth (Ph.D., Biochemistry)

The Human Microbiome and Interactions with Disease States, Medical Interventions, and the Metabolome

Thesis directed by Prof. Rob Knight

Next generation sequencing technologies have allowed researchers to evaluate the unculturable members of the human microbiome with unprecedented precision and detail. Early studies of the human microbiome revealed the unique microbial fingerprint that we contain, as well as the fact that numerous disease states, including allergies, obesity, and gastrointestinal disorders, are associated with alterations in our human microbiome or a lack of exposure to critical microbes. This dissertation examines some common disease states and their microbial associations.

Recent studies have also demonstrated how important it is to assess how medical interventions, including pharmaceutical drugs and implanted medical devices, alter our innate microbiome. Any negative alterations in microbial community structure or composition might have series deleterious consequences for the patient. Therefore, in this dissertation I assess the impact that intravaginal rings, used to deliver localized doses of acyclovir in herpes simplex positive women, have on the vaginal microbiome. I demonstrate that intravaginal rings are safe from a microbial perspective as they do not alter or aggravate the vaginal microbial community.

Furthermore, this dissertation discusses the importance of assessing our ancestral microbiome to determine how modern day society has changed our evolutionary microbiome, with possible harmful consequences. I discuss several mechanisms by which researchers are exploring how to replenish our natural defensive microbes. Along this line of research, I compare the microbial communities and functional profiles of the hunter-gatherer group Matses to individuals from the United States and those from the agrarian societies of Malawi and Venezuela. This work demonstrates how our ancestral microbiome was enriched in genes related to energy metabolism, and how altered our modern day microbial community structure is from our evolutionary past.

Finally, the importance of the interaction between intestinal metabolites and the microbiome is explored, focusing on how microbes can alter drug metabolism. I go on to evaluate the vast differences in the metabolite profile of potatoes that have been processed in different ways across a range of cultivar types. This work demonstrates that processing method is the dominant driver of metabolic profiles, but that cultivar and processing method interact non-linearly to produce unpredictable concentrations in the majority of potato-containing metabolites. I also show that common vitamins and antioxidants have unpredictable concentrations across processing methods and cultivars, highlighting that in the future this work might allow us manipulate our food metabolites to have maximum impact on the host.

## Acknowledgements

I would like to thank Rob Knight for his mentorship and introductions to fantastic collaborations, the members of the Knight lab for their fantastic help and friendship over the past few years, and my parents Ken and Cindy for their love and support.

CONTENTS

# TABLES

Table

# FIGURES

Figure

# Chapter I

## Introduction to the Human Microbiome

In this section I provide an introduction the human microbiome and the field of microbial ecology. Rapidly developing sequencing methods and analytical techniques are enhancing our ability to understand the human microbiome, and, indeed, how we define the microbiome and its constituents. This section highlights recent research that expands our ability to understand the human microbiome on different spatial and temporal scales, including daily timeseries datasets spanning months. Furthermore, emerging concepts related to defining operational taxonomic units, diversity indices, core versus transient microbiomes, and the possibility of enterotypes are discussed. The following text is modified from my reviews of the subject [1, 2].

### The Human Microbiome

The human microbiota consists of the 10-100 trillion symbiotic microbial cells harbored by each person, primarily bacteria in the gut; the human microbiome consists of the genes these cells harbor[3]. Microbiome projects worldwide have been launched with the goal of understanding the roles that these symbionts play and their impacts on human health[4, 5]. Just as the question, “what is it to be human?”, has troubled humans from the beginning of recorded history, the question, “what is the human microbiome?” has troubled researchers since the term

was coined by Joshua Lederberg in 2001[6]. Specifying the definition of the human microbiome has been complicated by confusion about terminology: for example, "microbiota" (the microbial taxa associated with humans) and "microbiome" (the catalog of these microbes and their genes) are often used interchangeably. In addition, the term "metagenomics" originally referred to shotgun characterization of total DNA, although now it is increasingly being applied to studies of marker genes such as the 16S rRNA gene. More fundamentally, however, new findings are leading us to question the concepts that are central to establishing the definition of the human microbiome, such as the stability of an individual's microbiome, the definition of the OTUs (Operational Taxonomic Units) that make up the microbiota, and whether a person has one microbiome or many. In this review, we cover progress towards defining the human microbiome in these different respects.

Studies of the diversity of the human microbiome started with Antonie van Leewenhoek, who, as early as the 1680s, had compared his oral and fecal microbiota. He noted the striking differences in microbes between these two habitats and also between samples from individuals in states of health and disease in both of these sites[7, 8]. Thus, studies of the profound differences in microbes at different body sites, and between health and disease, are as old as microbiology itself. What is new today is not the ability to observe these obvious differences, but rather the ability to use powerful molecular techniques to gain insight into why these differences exist, and to understand how we can affect transformations from one state to another.

Culture-independent methods for characterizing the microbiota, together with a molecular phylogenetic approach to organizing life's diversity, provided a fundamental breakthrough in allowing researchers to compare microbial communities across environments within a unified phylogenetic context (reviewed in [9]). Although host-associated microbes are presumably acquired from the environment, the composition of the mammalian microbiota, especially in the gut, is surprisingly different from free-living microbial communities[10]. In fact, an analysis of bacterial diversity from free-living communities in terrestrial, marine, and freshwater environments as well as communities associated with animals suggests that the vertebrate gut is an extreme environment[10]. In contrast, bacterial communities from environments typically considered extreme, such as acidic hot springs and hydrothermal vents, are similar to communities in many other environments[11]. This suggests that coevolution between vertebrates and their microbial consortia over hundreds of millions of years has selected for a specialized community of microbes that thrive in the gut's warm, eutrophic, and stable environment[10]. In the human gut and across human-associated habitats, bacteria comprise the bulk of the biomass and diversity, though archaea, eukaryotes, and viruses are also present in smaller numbers and should not be neglected[12, 13].

Interestingly, estimates of the human gene catalog and the diversity of the human genome pale in comparison to estimates of the diversity of the microbiome. For example, the Meta-HIT consortium reported a gene catalog of 3.3 million non-

redundant genes in the human gut microbiome alone[5], as compared to the ~22,000 genes present in the entire human genome[14]. Similarly, the diversity among the microbiome of individuals is immense compared to genomic variation: individual humans are about 99.9% identical to one another in terms of their host genome[15], but can be 80-90% different from one another in terms of the microbiome of their hand[16] or gut[17]. These findings suggest that employing the variation contained within the microbiome will be much more fruitful in personalized medicine, the use of an individual patient's genetic data to inform healthcare decisions, than approaches that target the relatively constant host genome.

Many fundamental questions about the human microbiome were difficult or impossible to address until recently. Some questions, such as the perennially popular "how many species live in a given body site?", are still hard to answer, due to problems with definitions of bacterial species and with the rate of sequencing error. Other questions, such as "how does the diversity within a person over time compare to the diversity between people?", or "how does the diversity between sites on the same person's body compare to the diversity between different people at the same site?", or "is there a core set of microbial species that we all share?", can now be answered conclusively. In the next section, we discuss some of the tools that have allowed these long-standing questions to be answered.

**Tools for Microbial Community Analysis**

The drastic reduction in sequencing costs experienced over the past few years

has made it possible to identify specific microbial taxa found within the human gut that are difficult or impossible to culture. Researchers are now able to generate millions of sequences per sample in order to assess differences in microbial communities between body sites and individuals. Our increased sequencing power has required the development of equally powerful computational tools to handle the burgeoning amount of sequence data produced by modern technologies[18]. There are several pipelines for analysis of microbial community data such as mothur[19], w.A.T.E.R.S[20], the RDP pyroseqeuncing tools[21], and QIIME (pronounced "chime")[22]. QIIME is a free, open-source platform for the analysis of high-throughput sequencing data that enables users to import raw sequence data and readily produce measures of inter- and intra-sample diversity. Consistency in the identification of operational taxonomic units (OTUs) and establishing agreed-upon measures of diversity within and between samples are crucial for the comparison of results across studies, although the concept of OTU is increasingly problematic as sequence data accumulate and explicitly phylogenetic approaches gain in popularity.

Beta diversity refers to the measurement of the degree of difference in community membership or structure between two samples. A recent review of taxon-based measurements of beta diversity found that some metrics, including Canberra and Gower distances, have increased power for discriminating clusters, while other metrics, such as chi-squared and Pearson correlation distances, are more appropriate for elucidating the effects of environmental gradients on

communities[23]. A robust method for comparing the differences between microbial communities is UniFrac, which measures the proportion of shared branch lengths on a phylogenetic tree between samples[24]. Highly similar microbial communities result in UniFrac scores near 0, while two completely independent communities that do not share any branch length (i.e. they have a different evolutionary history) would result in a UniFrac score of 1. Principal coordinates analysis (PCoA) can then visualize the Unifrac distances between samples in two-dimensional or three-dimensional space, allowing for the clustering of similar communities or separation of distinct communities to be easily distinguished visually.

UniFrac as a measure of beta diversity, coupled to PCoA, has the ability to distinguish differences between communities utilizing as little as 10 sequences per sample[25]. It is important to recognize that increased sequencing depth is not always necessary to recover biologically meaningful results when those results are obvious. Thus, by choosing diversity measurements that are appropriate for a study design, researchers utilizing modern sequencing methods are able to characterize differences between samples at relatively low sequence coverage. This enables researchers to assess fine-grained spatial and temporal patterns by characterizing hundreds to thousands of samples, such as timeseries across multiple patients or environments. The functionality of UniFrac, as well as a multitude of diversity measurements are available in QIIME and can be readily compared.

In general, pipelines for analyzing 16S rRNA and shotgun metagenomic data have separate workflows. Some initial steps, such as demultiplexing (removing

barcodes from and separating pooled samples) and quality filtering, are common to both pipelines. However, for 16S rRNA data, sequences must be grouped into OTUs, chimeric sequences generated by incomplete template extension must be removed, and phylogenetic trees must be constructed. In contrast, in the metagenomic pipeline, sequences must be assigned to functions as well as to taxonomy (either as whole reads or after assembly). Once taxon or gene function tables are constructed, the pipelines begin to converge, at least conceptually: the interest is then in 1) the composition of each sample, 2) finding the taxa or functions that discriminate among groups of samples (e.g. according to clinical parameters), and 3) in asking whether the samples cluster according to any measured clinical states (or according to time). One exciting emerging direction is comparing metagenomic and 16S rRNA clustering directly using a technique called Procrustes analysis that allows the PCoA plots to be combined[26]. Another powerful tool is the use of machine learning and statistical techniques to build predictive models of taxa[27] or functions[28] that discriminate between groups of samples.

A unique advantage of QIIME relative to other pipelines is its ability to exploit “sample metadata”, e.g. clinical information about subjects, to produce visualizations that make the main patterns in the data immediately apparent. Of particular interest, QIIME supports the MIMARKS (Minimum Information about a MARKer Sequence) standard[29] developed by the Genomic Standards Consortium[30], which is increasingly popular with other tools for microbial and community analysis such as MG-RAST[31], and has been adopted by the INSDC

(International Nucleotide Sequence Database Consortium, which includes GenBank, EBI, and DDBJ) as the standard for metadata.

With these tools in hand, basic patterns of similarities and differences in the microbiota are now routine. The key challenge now is to extend analyses to include longitudinal studies and to understand the role of specific host and environmental factors in the development and maintenance of the microbiome.

**Dynamic Interactions Between Human Microbes and the Environment**

The gastrointestinal (GI) tract of a human infant provides a brand new environment for microbial colonization[32]. Indeed, the microbiota that an infant begins to acquire depends strongly on mode of delivery[33]. Twenty minutes after birth, the microbiota of vaginally delivered infants resembles the microbiota of their mother's vagina, while infants delivered via Cesarean section harbor microbial communities typically found on human skin[34]. The acquisition of microbiota continues over the first few years of life, as an infant's GI tract microbiome begins to resemble that of an adult as early as 1 year of life[35]. In one case-study following an infant's microbiota over the first 2.5 years of life, phylogenetic diversity increases significantly and linearly with time[36]. Additionally, significant changes in gut microbiota composition were apparent at five time points; starting a diet of breast milk, development of fever at day 92, introduction of rice cereal at day 134, introduction of formula and table foods at day 161, and antibiotic treatment and adult diet at day 371[36]. Interestingly, each dietary change was accompanied by

changes in gut microbiota and the enrichment of corresponding genes. For example, as the infant began to receive a full adult diet, genes in the microbiome associated with vitamin biosynthesis and polysaccharide digestion became enriched[36].

The interaction between the human microbiota and the environment is dynamic, with human microbes flowing freely onto the surfaces we interact with everyday. Fierer et al. showed that human fingertips can transfer signature communities of microbes onto keyboards and these communities strongly differentiate individuals [37]. PCoA plots showed that it was possible to determine which fingers were typing on which keys, and which individuals were using which keyboards: it was even possible to link a person's hand to the computer mouse they use with up to 95% accuracy when compared to a database of other hands[37]. Overall, this study showed that microbial communities are constantly being transferred between surfaces, and that a dynamic interaction exists between environmental microbiota and different human body sites.

**Interpersonal Microbial Diversity**

Another interesting question that we are just beginning to answer is how stable the microbiome within an individual is over time. By defining what constitutes normal temporal variation in an individual over time, we will be better able to quantify and understand changes in microbial communities that result from dietary and pharmaceutical interventions. In the longest timeseries study to date, Caporaso et al. sampled two individual's microbial communities in the gut, oral

cavity, and left and right palms over 396 time points spanning 15 months[38]. Communities at different body sites were readily distinguishable from one another using 3-D PCoA plots over a one year time span, even though the community structure within a given site was highly variable[38]. The level of diversity is also different between body sites, with the mouth and gut harboring the most diverse communities[39]. Taken together, these studies show that an individual's microbiota represents a highly variable and compartmentalized ecosystem.

Overall, it has yet to be conclusively proven that individuals, or even body sites, harbor a "core" set of specific bacterial taxa. For example, the Meta-HIT consortium defined a "core" set of lineages as those that were present in half of the subjects studied, although essentially no genes were present in all subjects studied[5]. Of course, it is important to recognize that sampling depth may be critical for distinguishing taxa that are absent from those that are merely very rare; the dynamic range of microbial abundance is also quite large, and even within the Meta-HIT "core" genes, 2000-fold ranges of abundance were not uncommon. Proving that a taxon is completely absent in the gut is not possible with these types of studies, so core calculations should always carry with them a caveat about sequencing depth. Another factor to consider when defining diversity and a core is that methodological artifacts can greatly increase the apparent numbers of OTUs in a sample (and hence reduce the apparent fraction that is shared). Both sequencing error[40, 41] and issues related to alignment, especially multiple sequence alignment[42-45], can inflate the number of OTUs immensely. It is important to

ensure that the same methodological procedures were used when performing estimates of the core in terms of the fraction of individuals the core must be represented in, the minimum abundance, and the procedure for deciding which sequences count as "the same". Finally, there is a key question about whether variation around a core is structured so that humans harbor only a few general types of microbiota profiles in a given body site: this is well established for the vagina[46] but more controversial in the gut[47]. In general, extreme caution must be applied when performing clustering procedures, as many will break up continuous variation into clusters where none exist[23]. Robust model selection procedures that incorporate the possibility that only continuous variation, not discrete clusters, exist remain to be developed within the context of microbial community analysis.

There is increasing evidence that individuals actually share a "core microbiome" rather than "core microbiota". In a study of monozygotic and dizygotic twin pairs concordant for obesity or leanness, a subset of identifiable microbial genes, but not species, were shared between all individuals[17]. Remarkably, vastly different sets of microbial species yielded very similar functional KEGG pathways. However, deviations from this core microbiome were apparent in obese subjects, suggesting that it will be important to utilize metagenomic data in addition to determining microbial community composition with 16S marker gene studies when assessing differences between disease states. Understanding whether this principle holds true for other body sites will be fascinating; cross-biome metagenomic

comparisons have been exceedingly rare to date[48, 49].

**The Gut Microbiome Plays an Important Role in Digestion and Nutrition**

The evidence is mounting for the inextricable link between a host's microbiota, digestion, and metabolism. In an analysis of humans and 59 additional mammalian species, 16S rRNA sequences clustered together carnivores, omnivores, and herbivores in principal coordinate spacing, showing that community structures differ depending on diets[50]. Dietary changes in mice can also lead to significant changes in bacterial metabolism, especially small chain fatty acids and amino acids, in as little as one week[51], and can lead to large changes after only one day[52]. Importantly, the genetic diversity found within our gut microbiota allows us to digest compounds via metabolic pathways not explicitly coded for in the mammalian genome, greatly increasing our ability to extract energy from our diverse diets[53, 54].

Gut microbiota also seem to play an important role in obesity. Germ-free mice that receive a transplant of gut microbiota from conventional mice have an increase in adiposity without increasing food intake due to increased energy extraction from the diet and increased energy deposition into host adipocytes[55]. The two major microbial divisions, Firmicutes and Bacteriodetes, show different abundances depending on phenotype. Decreased Bacteriodetes and increased Firmicutes have been found in genetically obese mice (ob/ob) when compared to their lean counterparts[56], and the obesity phenotype can even be transferred to a

germ-free but genetically wild-type mouse by way of the microbiota, and the phenotype is due to energy balance: bomb calorimetry of the fecal pellets reveal that the ob/ob mice extract more energy from their diet, and leave less behind in the feces[53]. Fascinatingly, the same effects hold true for another mouse model, the TLR5 knockout mice, which also become obese in some mouse facilities (but develop colitis in others, presumably due to differences in the background microbiota). The TLR5 knockout mice also produce a transmissible obesity phenotype, but no difference in the efficiency of energy harvest is involved. Instead, the altered microbiota somehow makes the mice hungrier, and their microbe-induced obesity can be cured by restricting the amount of food in their cages to that consumed by wild-type mice, as well as by antibiotics[57]. The correlation between microbes and obesity is perhaps best illustrated through weight loss. As different groups of human subjects were placed on either a fat-restricted or carbohydrate-restricted diet, their abundance of Bacteriodetes increased as their body weight decreased, transitioning from the signature 'obese' microbial community to a 'lean' community[58]. Thus, the modulation of a patient's microbiota might be a therapeutic option for promoting weight loss in obese patients or promoting weight gain in underweight children.

Surprisingly, the microbes that we ingest with our food might be providing our individual microbiome with new genes to digest new foods. Hehemann et al. found that a new class of glycoside hydrolases used to digest porphyran, a polysaccharide common in red algae, was also found in human stool samples as a

gene in Bacteriodes plebeius. A closer examination of the stool metadata revealed that the stool samples containing the porphyran-digesting gene were only present in Japanese individuals; the gene was not found in the gut microbiome of the individuals of the United States. Why would a marine gene be found in human gut? The authors concluded that the seaweed common to the Japanese, but not American, diet contained the microorganism which transferred the genes to gut microbiome[59]. Thus, microbes have the ability to greatly increase the number of metabolic tools of the human gut, allowing us to digest an array of substrates.

**Plasticity of the Human Gut**

Given the relative stability of the human gut microbiota, one key question is whether it is sufficiently plastic to allow well-defined interventions to improve health. As described above, the gut microbiota is fairly stable over time once established, at least compared to the differences between individuals. However, a number of studies demonstrate that external forces can alter the community of microbes located in the GI tract and antibiotics are an important example.

Antibiotics are mainly used to combat pathogenic bacterial species that reside within or have invaded a host, however the current generation of antibiotics are broad spectrum and target broad swaths of the normal microbiota as well. Thus, antibiotics significantly affect the host's innate gut microbiota. Three to four days after treatment with the broad-spectrum antibiotic ciprofloxacin the gut microbiota experience a decrease in taxonomic richness, diversity, and evenness[60, 61]. The

large magnitude of changes in the gut microbiota demonstrated significant interpersonal variability. While the gut microbiota began to resemble it's pre-treatment state a week after treatment, differences between individuals were seen with regards to how closely the post-treatment community resembled the pre-treatment community, and some taxa failed to return to the community[61, 62]. Indeed, the reestablishment of some species can be affected for up to four years following antibiotic treatment[63]. Yet the overall recovery of the gut microbiota following antibiotic treatments suggests that there are factors within the community, biotic or abiotic, than promote community resilience, although these have yet to be elucidated.

Other antibiotics also tend to produce results that differ substantially between subjects[64, 65] and even body sites[66]. Because larger populations have not yet been studied, in part due to ethical issues with administration of antibiotics to healthy human subjects, the basis for these underlying differences has not yet been elucidated. Understanding the factors that determine the ability of a microbiota to resist and recover from perturbation, as well as understanding the factors that determine its current state, will be key to developing tools to assist in microbiome manipulation. For example, counter-intuitively, in rats the administration of antibiotics prior to cecal transplant actually reduces the chance that new microbes will establish[67].

One fascinating hint that the microbiota may be more plastic than imagined is the recent success of treatment of persistent Clostridum difficile infections via stool

transplant, which has been successful in a number of studies[68-74], and in general the depauperate gut community produced during the C. difficile infection is replaced by the donor community[69, 75]. The success of this technique is remarkable, especially considering how little is known about the best community to supply. For example, is it better to receive the fecal community of a close relative or of a cohabiting individual, or perhaps to bank one's own stool before beginning antibiotic treatment so that it can be restored later? Is the same stool good for everyone, or do the vast differences in the microbiota imply that each person's microbes are specifically adapted relative to those they might receive from a donor? As with blood types, are there "universal donors" and "universal recipients"? These and many other questions remain to be answered.

**Conclusion and Prospectus**

As in every year since the initial sequencing of DNA, this year has resulted in an unprecedented growth in the amount of sequence data collected at an unprecedentedly low cost. Increasingly powerful tools used to extract meaningful patterns from this wealth of data have been developed or updated as well. Emerging technologies such as stool transplantation, 16S rRNA and whole-genome sequencing on the Illumina platform, the ability to transplant human microbial communities into mice with high efficiency even from frozen samples[52], and the creation of personalized culture collections[76] raises the prospect of a future in which therapies for individual humans are piloted in a battery of mice that are subjected

to different treatments, and where leave-one-out experiments that reveal the effects of the deletion of individual species[76] or individual genes from within a species[77] allow insight into mechanism. Although the tools we have available are still imperfect (for example, the limited read length of today's high-throughput sequencing technologies limit the ability to detect bacterial species and strains, and analyses of viruses and eukaryotes are still very much an emerging frontier), the prospects for developing a mechanistic understanding of the factors that underlie the plasticity of the microbiome and then for manipulating the microbiome to improve health seem increasingly bright.

# Chapter II

## Intravaginal Medicated Rings and the Microbiome

This published chapter explores how alterations in the microbiome must be assessed when evaluating the efficacy of a medical device. Specifically, this chapter examines if the presence of an intravaginal ring for the localized delivery of acyclovir in women with herpes simplex virus alters their vaginal microbiome [78].

### Introduction

In 2003, an estimated 536 million people worldwide aged 15-49 were living with herpes simplex virus type 2 (HSV-2) with an annual incidence of 23.6 million [79]. Globally, HSV-2 is the most frequent cause of genital ulcer disease [80] and is associated with a three-fold increased risk for HIV-1 acquisition in women [81]. These epidemiological findings suggest that interventions against HSV-2 may have a key role in HIV prevention worldwide [82]. Daily oral valacyclovir (VACV), the 5'-L-valyl ester prodrug of the antiherpetic drug acyclovir (ACV), has been shown to prevent or delay genital recurrences by 85% [83] and to reduce the risk of transmission among HSV-2-discordant couples by 48% [80].

Topical application of ACV to the vagina is safe and has provided some clinical benefit for the treatment of primary or recurrent lesions by shortening their duration [84, 85]. We hypothesize that sustained delivery of ACV to the vaginal tract can provide an alternative approach to oral suppressive therapy and may

protect against sexual HSV acquisition. Delivering ACV from intravaginal rings (IVRs) holds potential benefits of improved adherence and low systemic exposure while maintaining steady-state levels in the vaginal tract. We previously developed a pod-IVR technology [86] that can deliver multiple compounds independently in a controlled, sustained fashion with pseudo-zero order kinetics [87-89]. The safety and pharmacokinetics of pod-IVRs delivering ACV in combination with the nucleoside reverse transcriptase inhibitor (NRTI) tenofovir (TFV) were evaluated successfully in the rabbit and sheep models [88]. We designed human silicone IVRs to release ACV and evaluated safety, pharmacokinetic, and surrogate efficacy in women with recurrent genital HSV, referenced herein as "genital herpes positive" (GHP), who switched their daily oral VACV suppression to the ACV IVR for 7 and 14 days [90]. This first-in-human study demonstrated that an IVR could safely deliver ACV and achieve comparable local mucosal levels to oral therapy without systemic absorption.

The vaginal microbial community has long been considered an important defense mechanism against infection [91-93]. Studies that sampled women across different ethnicities including Caucasian, African American, Hispanic, and Asian found that most vaginal communities could be defined by the presence of a dominating Lactobacillus species of *L. iners, L. crispatus, L. gasseri*, or *L. jensenii* [46, 94]. The other communities were not dominated by a Lactobacillus species, but still contained a dominant community of lactic-acid producing microbes [46]. The vaginal communities of Asian and Caucasian women were most often dominated by

lactic-acid producing Lactobacillus than Hispanic and African American women, possibly causing the lower vaginal pH levels found in Asian and Caucasian women. Bacterial vaginosis (BV) results in a significant community shift from healthy communities and negative health consequences [46]. Twenty nine percent of species-level OTUs were shared between healthy and BV-positive women, as BV-positive communities were characterized by decreases in Lactobacillus and increases in *Gardnerella, Atopobium, Megasphaera, Eggerthella, Aerococcus, Leptotrichia/Sneathia, Prevotella* and *Papillibacter* [95].

In studies on IVRs delivering TFV in pig-tailed macaques, we used confocal laser scanning microscopy, fluorescence in situ hybridization, and scanning electron microscopy to investigate IVR colonization by polymicrobial biofilms [96]. Large areas of the ring surfaces were covered with monolayers of epithelial cells that supported two biofilm phenotypes, both with a broad diversity of associated bacterial cells. Similar results were obtained in our clinical evaluation of IVRs delivering ACV in GHP women [90]. By Day 7, epithelial cell clusters had developed on the IVR surface, with little or no visible associated microbial growth. At Day 14, large areas of the ring surface were covered with a mat of epithelial cells that harbored the development of polymicrobial biofilms with similar morphological features to the biofilm phenotypes in our macaque studies.

Our limited understanding of how the vaginal microbiome responds to topical delivery of antiviral candidates is a critical gap in developing these strategies for clinical evaluation. Here we describe the first culture-independent assessment on

the bacterial colonization of IVRs in women and the concomitant effect on their vaginal microbiomes.

## Materials and Methods

*Participants and study design.* The participant characteristics and study design (**Figure 2.1**, **Table S2.1**) have been described in detail elsewhere [90]. Briefly, 6 HIV-negative, GHP women who were willing to change their suppressive oral VACV to an ACV IVR were enrolled into a pharmacokinetic and safety study. In order to prevent ACV washout from the vaginal tract, IVR insertion occurred within 24 h of oral VACV dosing, which was discontinued during the study. The first three participants used an ACV IVR for 7 days and had cervicovaginal lavage (CVL) collected prior to IVR placement; 1 and 3 days post-insertion; and at Day 7 when the IVR was removed. The final three participants used an ACV IVR for 14 days, and the study visits were extended to include sampling at 10 after IVR insertion and at Day 14 when the IVR was removed. The study design resulted in the collection of 30 CVL samples: 6 at IVR placement; 6 on Day 1; 6 on Day 3; 3 on Day 7; 3 on Day 10; 6 at IVR removal. None of the women displayed symptoms suggestive of active vaginal or sexually transmitted infection (STI) during the study.

*Sample processing and microbial DNA isolation.* ACV IVRs were removed aseptically on Day 7 or 14. The rings were cut into sections and portions of the segments without pods (i.e., unmedicated), were placed in 70% ethanol in water and

stored at 4°C. CVL samples were placed on ice and clarified by centrifugation at 700 × g for 10 min at 4°C. The cell pellets together with a fraction of their supernatants (ca. 1 mL) were stored at -80°C. Microbial DNA was isolated from a total of 36 samples (6 IVR and 30 CVL, see **Figure 2.1)** according to the following methods. Unmedicated IVR segments were cut into small pieces using a pre-sterilized scalpel and DNA was extracted from these samples using the PowerSoil® DNA Isolation Kit (MO BIO Laboratories, Inc., USA) according to manufacturer instructions. Frozen CVL samples were thawed on ice to afford a viscous fluid that was transferred into a microfuge tube and processed using the InstaGene Matrix (Bio-Rad Laboratories, USA) kit with the following modifications to the manufacturer instructions: sample volume to matrix volume ratio was 4:1 and the 56°C incubation time was extended to 40 minutes. Genomic DNA was isolated using two different methods, as the CVL samples contained inhibitors of PCR amplification that required a modified DNA extraction procedure.

Genomic microbial DNA isolated from the above samples was quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies Corporation, USA) according to the manufacturer's instructions. The isolation procedures yielded between 9.7 and 22.5 μg (CVL, mean 15.1 μg) and between 3.3 and 35.5 μg (IVR, mean 14.5 μg) of high quality, PCR inhibitor-free, whole genomic DNA per vaginal sample.

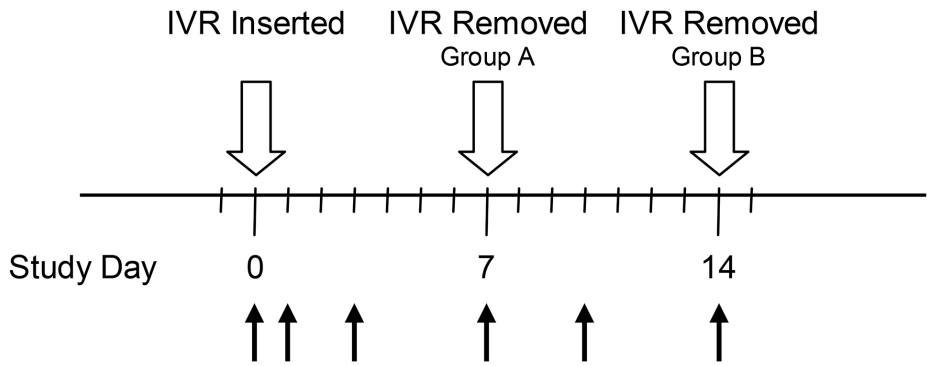


**Figure 2.1**. **Study timelines and CVL sample collection points** (black arrows). Participants used an ACV IVR from Day 0 (24 h after last dose of oral VACV) to Day 7 (Group A, n = 3) and Day 14 (Group B, n = 3). CVL was collected for microbial DNA isolation on: Group A, Days 0 (pre-dose), 1, 3, and 7 (upon IVR removal); Group B, Days 0 (pre-dose), 1, 3, 7, 10, and 14 (upon IVR removal).

*DNA amplification and sequencing of 16S rRNA genes*. Amplification and sequencing of the V4 hypervariable region of the 16S rRNA gene was performed using the validated, region-specific bacterial/archaeal primers 515F and 806R according to previously described methods [97] optimized for the Illumina MiSeq platform. 5'-Barcoded amplicons were generated in duplicate using Premix Ex Taq (TaKaRa, Japan) and a MyCycler thermal cycler (Bio-Rad Laboratories, USA). The PCR conditions consisted of an initial denaturing step of 94°C for 2 min, followed by 7 cycles of 94°C for 30 s, 48°C for 30 s, and 72°C for 1 min, 28 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 1 min, and a final elongation step of 72°C for 10 min. Replicate reactions were pooled and the amplicons were separated by electrophoresis in 1.0% agarose gels. The amplicons were purified using QIAquick Gel Extraction kit (Qiagen, USA) according to manufacturer's instructions. The $A_{260}$:$A_{280}$ absorbance ratio was acquired with a SpectraMax® Plus Absorbance Microplate Reader (Molecular Devices, USA) and used as an indicator of DNA purity. Amplicon DNA was quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies Corporation, USA). Amplicon aliquots (100 ng) from all 36 samples were pooled and re-purified with the UltraClean® PCR Clean-Up Kit (MO BIO Laboratories, Inc., USA). The purified, pooled sample contained 8.8 µg of DNA with an $A_{260}$:$A_{280}$ ratio of 1.82 and was submitted for sequencing using the MiSeq platform (Illumina Cambridge Ltd., United Kingdom) at the Advanced Genomics Facility, University of Colorado, Boulder, USA. The sequence data will be submitted to the European Molecular Biology Laboratory European Bioinformatics

Institute at the time of publication.

*Microbial community analysis*. The 16S rRNA sequences obtained from the MiSeq platform were processed through the open source software pipeline Quantitative Insights Into Microbial Ecology (QIIME) version 1.7.0 [98]. Sequences were filtered for quality using established guidelines [99]. Quality-filtered reads were then demultiplexed, yielding 3,989,623 sequences total with an average length of 151 bases per read, and an average coverage of 110,822 sequences per sample. Sequences were then binned into Operational Taxonomic Units (OTUs) based on 97% identity using UCLUST [100] against the Greengenes reference database [101] May 2013 release. The representative sequences for each OTU were then compared against the Greengenes database for taxonomy assignment. Each sample's sequences was rarefied to a depth of 28,000 sequences per sample to reduce the effect of sequencing depth, and used for downstream analysis. This level of rarefaction was chosen to minimize the number of samples dropped from downstream analysis but maximize the number of sequences allowed per sample. Following rarefaction, 1 sample (Subject60.CVL14.671842) was omitted from further analysis due to insufficient coverage, yielding 35 samples and 581 OTUs in 980,000 sequences.

A Poisson embedding algorithm was employed to determine a 95% confidence interval for finding a new OTU [102] and, consequently, to calculate the proportion of the overall community diversity captured by sequencing. The overall highest probability for discovering new OTUs in our samples was 1.295%, meaning that our

sequencing had revealed 98.705% of the overall community diversity (**Table S2.2**). The β diversity of samples was measured using the weighted UniFrac metric [103], and the dimensionality reduction technique of Principal Coordinates Analysis (PCoA) was used to visualize the community differences. The statistical cutoff of P = 0.05 after False Discovery Rate (FDR) correction for multiple comparisons was used to define statistical significance when testing if taxa were significantly different between groups.

**Results**

*Sample size considerations*. Demonstration of safety is a central component in the development of topical microbicides for the prevention of sexual HIV transmission and typically has involved evaluating general toxicology and irritation/inflammation in one or more animal species [104] along with a rudimentary (culture-dependent) assessment of the vaginal microflora in macaques [105, 106]. The application of a culture-independent analysis to determining the effect of the microbicide IVR on the vaginal microbiota therefore was performed during early safety evaluation of this novel device, using six subjects in an exploratory clinical trial. Clinical trial failures with Nonoxynol-9 (N-9), Carraguard, C31G (Savvy), and cellulose sulfate (CS) underscore the urgent need for comprehensive, early microbicide efficacy and safety assessment prior to large, Phase II/III clinical trials [107]. This sample size was able to capture the nearly all of the beta diversity of the vaginal community found in the mid-vagina samples

obtained from the HMP (**Figure S2.3**) using the weighted UniFrac distance metric, as 20 samples at 1000 sequences per sample explained more than 99.5% of total beta diversity. Given the novelty of this report, there are few relevant precedents that can be used to make group size comparisons with. Ravel et al. used 16S rRNA gene pyrosequencing to characterize the vaginal microbiota in samples from 35 healthy women in a two-week study of twice-daily application of 1 of 3 vaginal gel formulations: placebo (10 subjects), 6% CS (13 subjects), and 4% N-9 (12 subjects) [108]. Despite the small cohort sizes, an inter-group comparison was possible and found that treatment with active microbicides shifted the microbiota toward a community type dominated by strict anaerobes and lacking significant numbers of Lactobacillus spp. These results support the cohort size used here for inter-group comparison, requiring less statistical power.

*Classification of vaginal microbiota.* Ravel et al. have shown that the vaginal bacterial communities of 396 asymptomatic women could be broadly classified into 5 major community state types (CSTs) [109]. Two of the women in the current study had vaginal microbiota dominated by Lactobacillus iners, one dominated by L. helveticus, and the remaining 3 woman had a more diverse vaginal community, consisting mainly of Lactobacillus, Proteobacteria, the Bacteroidetes genus Prevotella, and the Actinobacteria Gardnerella vaginalis in varying proportions (**Figure S2.1 and Figure 2.2**). The vaginal bacterial communities of these 5 women are consistent with 2 of the 5 proposed major CSTs (**Figure S2.1**). One of the study participants, however, had a vaginal community that was dominated

(91% relative abundance) by Lactobacillus helveticus, a community structure that does not match with any of the CSTs. Furthermore, our analysis suggests that L. crispatus was not present in any of the samples although this species makes up a dominant member of one of the CSTs [109]. Larger cohorts of GHP women will be needed in future analyses to determine if these taxon differences are a consequence of GHP status, VACV suppressive therapy, or merely reflect the small sample size. The vaginal bacterial communities of the 6 study participants cluster primarily by individual (**Figure 2.2**), an expected result given the large interpersonal variation observed across body sites [110].

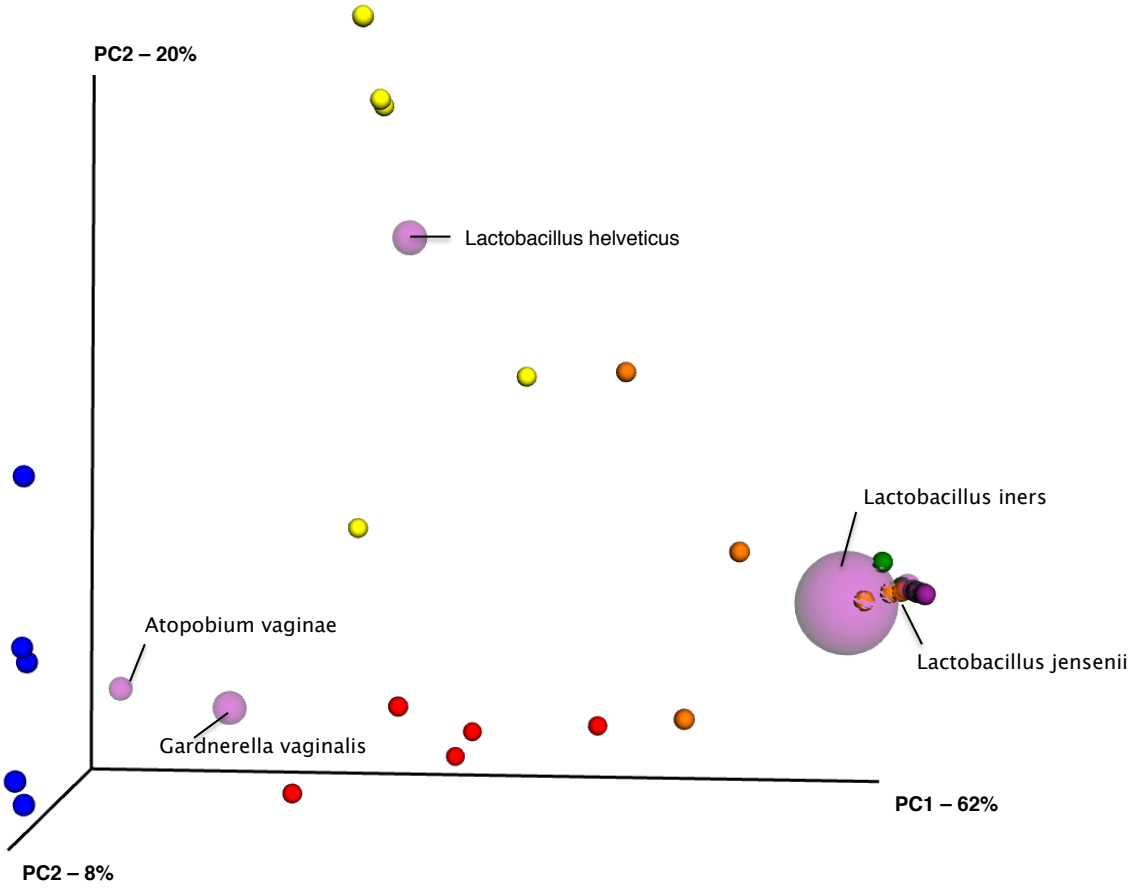


**Figure 2.2. Comparison of the vaginal bacterial communities among GHP women.** Weighted UniFrac distances plotted in PCoA space of vaginal microbial communities from the 6 GHP study participants, each represented by a distinct primary color, reveals clustering by individual. The labeled purple spheres represent taxa, and the proximity of colored participant samples to the taxa spheres is indicative of increased membership of that taxon in a given community. The larger the grey taxon sphere, the greater that taxon's overall abundance. Three patients had communities dominated by Lactobacillus iners , one patient was dominated by L. helveticus, and two other patients had communities with a shared proportion of Atopobium vaginae and Gardnerella vaginalis in addition to Lactobacillus.

*Duration of GHP status is correlated with an altered vaginal microbiome.* A longitudinal analysis was conducted to investigate associations between the abundance of bacterial taxa and the duration of GHP status (median: 4.0 years, 2-8 year range) in the 6 study participants (**Table S2.1**)A significant correlation was found between the abundance of Pseudomonas spp. (correlation ~ 0.45, P ~ 0.05) and Gardnerella vaginalis (correlation ~ 0.68, P ~ $3.6\times10^{-4}$) with the duration of GHP status in these samples. These results are in agreement with previous reports that identified an association between GHP women and an altered vaginal microbiota [111, 112]. We observed a significant difference in the α diversity in vaginal samples from women with duration of GHP status of 2 and 7 years (**Figure 2.3**) (4.2× more OTUs, P <<0.001; phylogenetic diversity metric PD_whole_tree 2.2x higher, P << 0.001). The vaginal bacterial communities of healthy women commonly have a low α diversity and are dominated by a smaller number of Lactobacillus spp. [109]. The altered vaginal community structure of GHP women, shown here to be correlated with duration of GHP status, is known to be associated with increased incidence [111, 112] and recalcitrance [113] of bacterial vaginosis (BV), as well as an enhanced susceptibility to an overlapping spectrum of other bacterial STIs [111]. An correlation also was identified between the number of genital HSV recurrences per year and a group of 6 bacterial taxa, consisting of members of: family Veillonellaceae (1 OTU, P = $2.74\times10^{-16}$), genus Dialister (1, P = $8.29\times10^{-13}$), genus Prevotella (3 OTUs, P-values $8.81\times10^{-12}$, $5.89\times10^{-4}$, and $4.99\times10^{-4}$), and genus Anaerococcus (1 OTU, P = $7.59\times10^{-4}$).

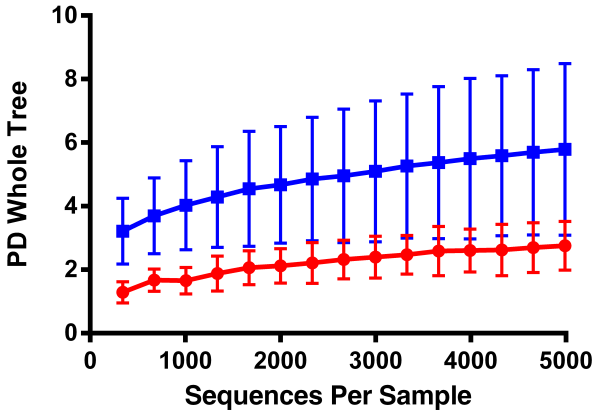


**Figure 2.3**. **Comparison of α-diversity in vaginal microbiota based on length of GHP status.** Subject with recurrent GHP status of 7 years (blue trace) has significantly higher α diversity, as measured with the phylogenetic diversity metric PD_whole_tree, compared to a subject with GHP status of 2 years (red trace).

*Microbial biofilms developing on the IVRs consist of communities reflective of those in the vaginal tract.* At Days 7 and 14, the IVR surface was covered with microbial biofilms at different stages of development [90]. The structure of these sessile communities was found to be representative of the corresponding communities established on the hosts' vaginal epithelium. A weighted UniFrac β diversity plot showed an overlap of vaginal (CVL) and IVR samples, and taxonomic comparison revealed that no taxa were significantly different between these data sets across all time points (**Figure 2.4**). Comparing the microbial communities developing on the IVRs (Day 7 or Day 14) with the corresponding vaginal microbiota in samples on the day of IVR removal did not identify any taxa that were different across all participants, suggesting that the IVR biofilms were not systematically enriched with any members of the vaginal microbiome. There were no significant taxonomic differences in the microbial communities between the IVR biofilms that were removed on Day 7 compared with those removed on Day 14, nor was the α diversity of the IVR samples different from the vaginal (CVL) samples (**Figure S2.2**).

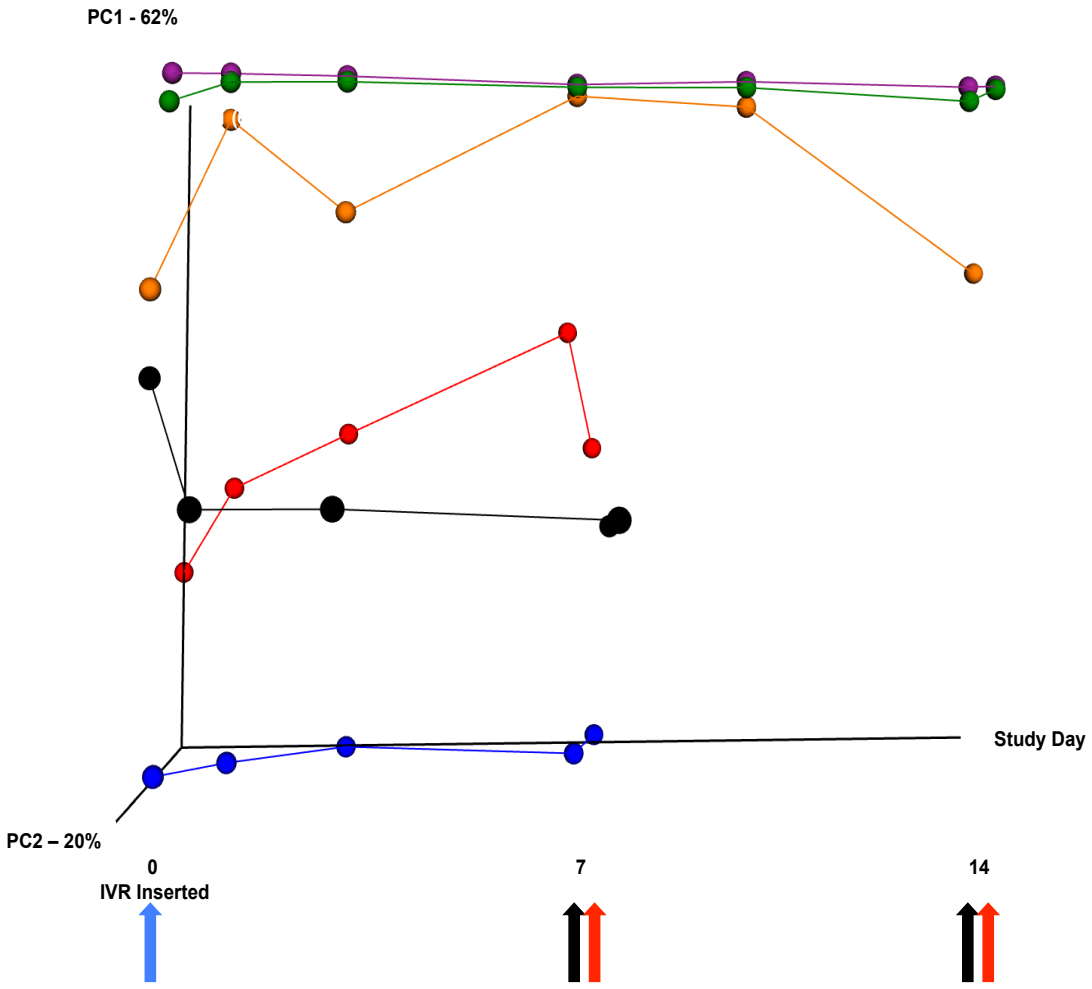


**Figure 2.4**. **Temporal dynamics of the vaginal microbial communities**. The β-diversity for CVL and IVR samples from the 6 subjects is plotted against time. Each colored tracing line, and correspondingly colored points, represents one subject who had the IVR removed either at Day 7 or Day 14 (see Fig. 1). The blue arrow points to CVL samples taken at the time of IVR insertion. The black arrows represent the CVL samples taken at the time of IVR removal (red arrows). The patient represented by the orange line and spheres did not have a CVL sample that met minimum 16S processing standards and was not included in the plot.

*IVRs do not significantly alter the hosts' vaginal communities over time.* Comparing the vaginal (CVL) and IVR samples on Days 7 and 14 assessed the effect of IVR administration length on community composition. Vaginal samples from participants who received the IVR for 14 days contained significantly fewer members from the phyla Proteobacteria and Firmicutes and from the genera Bacillus and Sphingomonas compared to participants who received the IVR for 7 days (relative abundance $3.58\times10^{-05}$ versus 0, $P = 9.91\times10^{-07}$; relative abundance $3.58\times10^{-05}$ versus 0, $P = 4.96\times10^{-07}$). Additionally, there were some subject-specific differences in the abundance of select taxa when comparing the IVR biofilm microbiota to the corresponding vaginal community, but no taxa were significantly enriched in the IVR biofilms across all samples (**Table 2.1**). The observed differences in vaginal microbiota were reflective of the large interpersonal differences, not driven by IVR usage.

| Taxon | P-Value | Average relative abundance (%) | |
|---|---|---|---|
| | | Healthy women, HMP [117] | GHP women, this study |
| Lactobacillus helveticus | $9.01\times10^{-13}$ | 0.2 | 13.0 |
| Gardnerella vaginalis | $2.89\times10^{-9}$ | 2.1 | 12.0 |
| L. iners | $2.00\times10^{-6}$ | 18.9 | 48.6 |
| L. crispatus | $1.19\times10^{-5}$ | 0.1 | 0.0 |
| L. reuteri | $6.56\times10^{-3}$ | 0.7 | 0.0 |
| L. intestinalis | $19.7\times10^{-3}$ | 2.0 | 0.0 |

**Table 2.1. Summary of taxa differences between vaginal communities of healthy women from the HMP and women with a history of recurrent genital herpes**. Each sample was rarefied to 2,000 sequences per sample. ANOVA was used to calculate if the relative abundance of any taxa were significantly different between the two groups. The P-values reported above are after FDR correction.

Studies on the dynamics of the human vaginal microbiota have shown that some communities can shift over short time periods while others remain relatively stable [114, 115]. In order to determine the effect of the IVR's presence on the temporal dynamics of the vaginal communities, the β diversity in each participant's IVR and vaginal samples was plotted over time (**Figure 2.4**). The CVL and IVR samples were connected by a solid line through PCoA space, with the first point representing the starting CVL microbial community prior to IVR administration, and the final point representing the IVR microbial community. Some of these microbial communities remained stable over the study period, while other changed markedly over time. However, no taxa were significantly different in the vaginal samples collected prior to IVR administration compared with the corresponding samples collected when the IVR was removed, suggesting that the presence of the IVR is not selecting for any specific taxa across all subjects. These findings will need to be confirmed in future studies using larger cohorts to detect more subtle differences.

**Discussion**

Culture-based evaluation of IVR effects on the vaginal microbiota has played an important role in early clinical assessment of candidate products for 40 years. In 1973, Henzl et al. evaluated early reservoir-IVRs delivering the hormonal contraceptive chlormadinone acetate in 12 women over 3 consecutive menstrual cycles [116]. The microbiologic examinations of the vaginal secretions consisted of direct smears stained with Gram's stain and a special stain for Trichomonas

vaginalis. In addition, the secretions were evaluated by cultivation on broad-spectrum media for aerobic and anaerobic microbes and a selective media for mycoses, Trichomonas vaginalis, and Neiserria spp. In the early 1980s, Population Council reservoir-IVRs [118] delivering the contraceptive levonorgestrel (LNG) in combination with estradiol (E2) were studied for contraceptive effectiveness and acceptability in multicentered trials involving 1,103 ring users [119]. Increased vaginal discharge, potentially due to alteration of the vaginal microbiome, was identified as the most frequently voiced complaint by users [119, 120]. This possibility was disproven by Schwan and colleagues in a group of 17 subjects who used the Population Council LNG-E2 IVR for 6 months [121]. Culture-dependent methods were used to characterize the aerobic and anaerobic bacteria, yeast, and mycoplasma/ureaplasma in vaginal secretions collected from the cervix and posterior fornix before and after treatment with the IVR. No significant difference was observed between both groups and it was concluded that alterations in the vaginal bacterial ecology through IVR usage was not the cause of the increased discharge.

While culture-based methods have continued to be used in clinical evaluation of IVRs [122], the recent introduction of high-throughput, culture-independent molecular methods [97] has made it possible to collect hundreds of thousands of sequences spanning hundreds of samples. The so-called democratization of sequencing [123] has continued to be fueled by ever decreasing sequencing costs, notably on several Illumina platforms [97]. The Illumina sequencing platform

returns on the order of 100 million sequencing reads per flowcell and therefore supports unprecedented sequencing depth [124] enabling the detection of very rare phylotypes [125]. Deep sequencing methods allow important parameters for describing microbial community composition to be determined, including the species richness within each sample (α diversity), and the diversity shared between multiple environments (β diversity) [126, 127]. We have used bar-coded Illumina sequence datasets generated from 16S rRNA gene fragments to study the response of the vaginal microbiomes of GHP women to pod-IVRs medicated with ACV. This rigorous analysis allowed the vaginal communities developing in microbial biofilms on the IVR surface to be compared for the first time to the corresponding vaginal microbiomes. Our results suggest that rigorous, molecular analysis of the effects of intravaginal devices on the corresponding microbial communities shows promise for integration with traditional approaches in the clinical evaluation of candidate products.

**Conclusion**

Despite the small sample size (n = 6), the study supports the preliminary safety of the ACV pod-IVR as there were no detectable changes in the vaginal microbiome in response to the devices. Although microbial biofilms were readily detected on the IVR surface, the composition of these sessile communities was similar to that of the corresponding vaginal microbiome. Notably, differences were observed in the vaginal microbial communities in these six GHP women compared

to healthy women participating in the HMP. Specifically, none of the GHP women had a L. crispatus dominant community and the GHP group tended to have increased numbers of G. vaginalis and L. iners. These findings are consistent with other studies suggesting a link between HSV-2 and BV [111, 112]. Prospective longitudinal studies using molecular analyses of the vaginal microbiome are needed to determine if HSV-2 triggers changes in the microbiome or, conversely, if changes in the microbiome promote HSV-2 reactivation. The observation that microbial α diversity was impacted by the duration of GHP status suggests that HSV-2 may promote changes in the microbiome.

# Chapter III

## Replenishing our Defensive Microbes

This published chapter introduces the importance of assessing our ancestral microbiome, how modern day changes in lifestyle could be altering our microbiome and contributing to disease states, and methods for replenishing our defensive microbes [128].

### Introduction

The relatively recent transition of human populations from hunter-gatherer and agricultural societies to industrialized societies has been concurrent with a rise in previously absent "Western" diseases, including obesity, asthma, and inflammatory bowel disease. The 'hygiene hypothesis' is one of many hypotheses proposed to explain this increase. This hypothesis suggests that industrial societies reduce our exposure to microbes with which we have coevolved, leading to improper immune function and to establishment of microbial communities that differ substantially from those of our ancestors [129].

Although the roles of specific pathogens have received intense scrutiny, we have only recently begun to understand the importance of microbes that can positively influence human health. The rapidly decreasing costs of DNA sequencing now allowing analyses of the microbes that live in and on the human body on a scale and with a resolution that has not previously been attainable. Large consortia, such

as the NIH's Human Microbiome Project [130] and the EU's MetaHIT [131], together with many other human microbiome projects on different scales worldwide, have given us a first impression of the diversity of the human microbiome. These projects allow us understand the microbes that we harbor, how these microbes assemble into healthy communities, and the genes involved in specific microbial functions. In this review, we discuss the mechanisms by which our microbes train our immune system to recognize and overcome pathogens, how modern societal practices may derail our microbiomes from their ancestral tracks, and how replenishing our microbiota with beneficial microbes can improve human health.

**Microbes Provide Health Benefits and Protection Against Pathogens**

The modern infatuation with cleanliness stems in part from the misguided midcentury thinking that most microbes cause disease, and that the absence of microbes is therefore a key component of health. Over the last twenty years, the use of culture-independent methods that allow us to identify the members of human-associated microbial communities that are difficult to grow in the laboratory, together with epidemiological studies and studies of germ-free mice, has started to change this thinking. There is now compelling evidence that the opposite is true: rather than reducing microbial exposure, we should balance our symbiotic microbial communities to protect us from pathogens and disease states. The specific features of the microbial communities that provide protection varies considerably, and depends on what is being protected against. In some cases, such as [132], a single

microbial species can provide protection; in others, such as [133, 134], half a dozen specific members are required; and yet others, such as [135, 136], require the action of a much larger community in aggregate. Although specific roles played by some important microbes have been identified, as outlined below, the full range of protective effects and their causative microbial agents remains unknown. In this review, we focus on bacteria, but it is important to recognize that viruses, eukaryotes, and even archaea are also important member of the human microbiota, and their effects on health are also important and increasingly studied [137, 138]. Additional discoveries about the interplay between the host GI tract, immune system, and environmental microbial communities continue to accumulate at a rapid pace [139]. Here we focus on two of the best-established ways in which microbes contribute to resistance against pathogens and pathogenesis.

**Microbes 'Educate' the Immune System Through Direct Interaction, Which is Required for Proper Immune System Response**

The reaction of the human immune system to both constitutive and transient members of our microbiota requires a delicate balancing act. The immune system must produce enough pro-inflammatory signals to recruit, differentiate, and cause the proliferation of effector cells to control populations and localization, but must also avoid inflammatory responses that would damage the host [140]. Because both beneficial and harmful microbes exist, our immune system must be 'taught' to identify and respond to each microbe correctly. This education of the immune

system requires direct contact with the microbiota (for example, by antigen-presenting cells and toll-like receptor signaling), and alters key components of both the adaptive and innate response. In the adaptive immune system, studies in germ free (GF) mice show that colonization of the gut is critical to the induction of regulatory T cell ($T_{reg}$) populations [141]. These induced regulatory T cells ($iT_{reg}s$) promote gut health by balancing the pro-inflammatory response with an anti-inflammatory one. Additionally, when microbes pierce the mucosal or epithelial barrier, they respond by releasing anti-inflammatory cytokines that reduce the intensity of the Th2 skewed response [142]. In mouse colitis models, transfer of naive T cells and $iT_{reg}s$ can both ameliorate symptoms and prevent development of symptoms in mice that are genetically predisposed to develop colitis [143]. Different microbes induce different $iT_{reg}s$, and the community of $iT_{reg}s$ is therefore influenced by historical and ongoing microbial exposure [144]. For example, Bacteroides fragilis induces $T_{reg}s$ via secretion of polysaccharide A [145]. This induction can ameliorate colitis symptoms, but depends on induction of the correct $iT_{reg}s$. Similarly, certain species of Clostridium induce $T_{reg}s$ that prevent or reduce colitis in mouse models [146], although the specific mechanisms by which they trigger induction are not well understood. Prior exposure to Clostridium species can also be important for the induction of $T_{reg}s$ in the context of Helicobacter pylori infection, determining the severity of infection [147]. Because the $iT_{reg}$ population is exposure-dependent, the $iT_{reg}$ repertoire provides a mechanism by which past disruption of the gut ecosystem might cause later dysbiotic or

pathogenic events. Shifts in the microbial population away from the $iT_{reg}$ repertoire might decrease the anti-inflammatory capacity of the gut immune system, creating an aggressive response that can alter gut microbial community structure and/or cause host tissue damage.

Acute infection has long been known to transiently change the states of many components of the immune system, not just $iT_{regs,}$ but we are only now beginning to realize how nonpathogenic members of our gut microbiota alter the state of the immune system over longer timescales (for example, by affecting antibody and defensin production). GF mice without diverse microbial communities cannot produce normal levels of antibodies upon inoculation with pathogens [148]. Similarly, antibody responses to viral infection in the lung mucosa depend on specific commensal microbes [149]. Although the mechanisms by which these microbes alter the state of the immune system are not yet well understood, commensal microbes appear to prime the immune system outside the context of infectious disease. Thus the diversity of the microbiome, and the past microbiome of an individual, might be critical components of health.

## Biodiversity of Microbial Communities Plays an Important Role in Preventing Disease and Infection

Ecological studies on larger scales suggest more diverse communities are in general more robust to invasion or disruptive events [135]. Biodiversity can also limit the emergence and spread of disease, in part through changes to the

community structure that are not possible in less complex communities [150]. On the smaller scale of human-associated microbial communities, diversity may play a similar protective role. A diverse microbiome might provide protection by many different mechanisms. Some of the best-supported hypotheses about diversity-protection relationships are that diverse communities might: use resources that would otherwise be available to a pathogen [17], produce short chain fatty acids such as butyrate or other molecules that inhibit growth of pathogens [151], or directly modulate the immune system effector population and/or cytokine milieu [152, 153]. However, support for all of these hypotheses is limited, and it is not yet even known in general whether the protective effect of biodiversity is a community-level effect, or whether high levels of biodiversity simply increase the probability that a particular species that is protective against the condition of interest is included in the community. Three specific cases are intriguing. First, in the locust gut, pathogen invasion was limited by overall community diversity, and not by the presence of any specific member [135]. Second, transferring the microbiota between strains of mice (NIH to C3H/HeJ) eliminated the susceptibility of C3H/HeJ mice to Citrobacter rodentium infection [133]. However, in contrast, C. difficile infection susceptibility could be altered by introducing only a small subset of a resistant host's community, suggesting that only a few specific members were involved in this case. These results are not necessarily contradictory -- some disease states or susceptibilities could stem from low biodiversity, others from the absence of a specific microbe. Further research could investigate whether antibiotics have

deleterious effects proportional to the extent to which they chronically reduce bacterial diversity in the gut. Hypertrophic environments might also reduce the benefit of the endogenous gut microbiota by preventing them from scavenging the majority of available resources, thus allowing a 'weedy' or pathogenic species to establish and expand itself and reducing overall diversity (as is seen in other hypertrophic environments on other scales) [17].

**Manipulation of the Microbiota is a Promising Method for Treating Disease**

Because the gut microbiota activate host immune defenses that are critical for protection against infection, microbiome manipulation is developing as an increasingly important treatment modality. For example, as noted above, NIH Swiss mice can resolve colonization with the murine pathogen C. rodentium, but the same infection in C3H/HeJ mice is lethal. However, transferring gut microbiota from NIH mice into C3H/HeJ mice delayed pathogen infection and mortality. These improvements were associated with increases in IL-22 in mice that received the NIH mouse microbiota transplant, suggesting that the microbiota is fine-tuning the host's innate immune system to prevent infection [133]. More work is needed to understand what role the microbiome plays in human diseases, and especially how many other effects attributed to host genetics may actually stem from shared vertically transmitted microbial species or communities.

The skin microbiome also plays a critical role in defense against pathogens.

The skin acts as the body's first line of defense against incoming pathogens, and production of effector T cells is linked to signals produced by host-associated microbiota. For example, GF mice monocolonized on the skin with Staphylococcus epidermidis produced significantly more proinflammatory IL-17A in the skin, but not in the gut, than did uncolonized mice [154]. Effector T cell production and function in the skin was unaffected by antibiotics that substantially changed the gut microbiota, suggesting that these two reservoirs of microbes modulate host immunity independently. Staphylococcus epidermidis also reduced dermal infection by the parasite Leishmania major, primarily by augmenting IL-1 cell signaling to activate local effect cell responses [154]. These studies demonstrate how the microbial communities at different body sites can be protective, and additional studies of the nares, vagina, mouth, and other body habitats are likely to extend these results to other body sites.

**Modern Behaviors Reduce our Exposure to Possible Beneficial Microbes**

Studies of the gut microbiome of modern humans living in remote, traditional communities, and of ancient humans from fossil or subfossil specimens, are beginning to provide a foundation for understanding how modern, "Westernized" humans have altered their gut microbiome from ancestral states [139, 155-157]. Humans in rural, remote Malawian and Venezuelan communities have differ markedly in their gut microbiota and microbiomes from humans living in the highly westernized US. Some of these differences may have evolutionary roots. For

example, 1400-year-old human fecal material from a high-altitude rock shelter in El Zape, Mexico preserved a gut microbiome that resembled the microbiome of humans currently living in Malawi and Venezuela, and differed from the microbiome of individuals living in the US [157] (**Figure 1**). For example, the spirochaete Treponema berlinense was found both in ancient El Zape fecal material and in rural, traditional populations of Malawi and Venezuela, but not in the United States population. As studies expand to include more human populations living traditional lifestyles and/or additional sources of ancient samples, general patterns and associations may allow us to characterize the pre-antibiotic, ancestral state of the human gut.

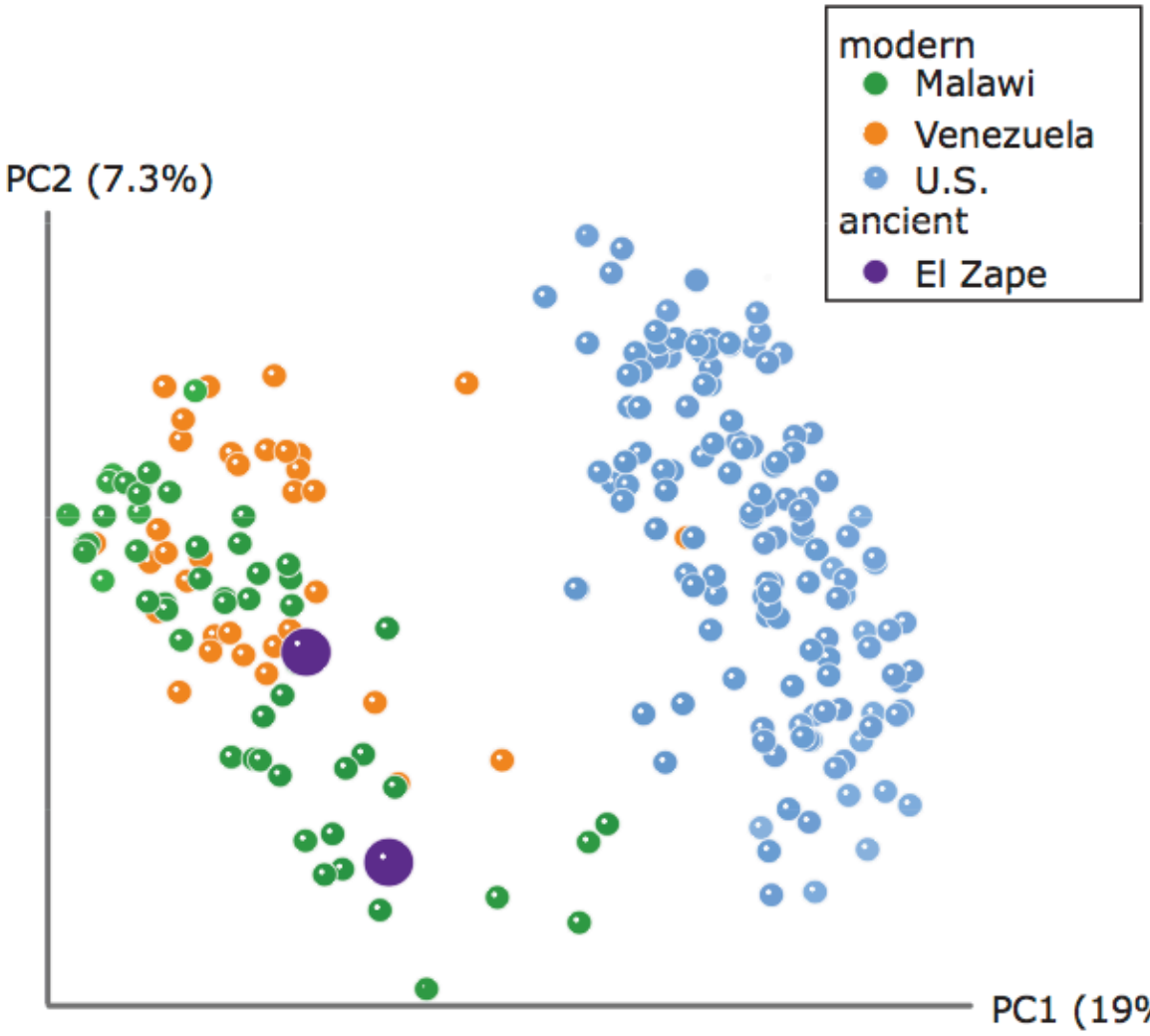


**Figure 3.1. 16S rRNA gene sequencing survey reveals major differences in community composition of ancient vs. present-day humans**. Fecal samples were collected from four different sources; adults in present-day Malawi, adults in present-day Venezuela, adults in present-day United States, and a ~1400 year old deposit in an ancient rock shelter in El Zape, Mexico. Briefly, the 16S rRNA gene was sequenced, taxonomy was assigned against a reference database, and the communities were compared using the unweighted UniFrac phylogenetic distance metric. Microbial communities that are more dissimilar are located further apart in principal coordinate space, while similar communities are found clustered together [155, 157].

These studies may even provide a pool of possibly beneficial ancestral microbes that have been lost due to recent lifestyle changes and that could be resupplied to improve health. We describe how several aspects of westernized societies - antibiotics, Cesarean sections, and lack of exposure to livestock - may be significant drivers of microbial change.

**Antibiotics**

Many studies have shown that antibiotic use in humans drastically decreases gut microbial diversity [61, 62, 158]. Although antibiotics are immensely valuable for clearing life-threatening infections, their overuse in patients may lead to unintended consequences. As noted above, a diverse gut microbiota can be protective against disease, and increasing evidence suggests that the depletion of this diversity by antibiotics may increase susceptibility to later infections. For instance, mice dosed with the antibiotic ampicillin were much less resistant to colonization when dosed with $10^8$ CFUs of vancomycin-resistant Enterococcus faecium (VRE) than controls that did not receive antibiotics. The gut communities of the antibiotic-treated mice were completely dominated by VRE. Remarkably, the gut microbial communities of humans receiving the same antibiotics were dominated (>97%) by the genus Enterococcus just 7-18 days prior to VRE infection in the bloodstream, demonstrating that antibiotic use might reduce the community's ability to fight off invading microbes [159].

Repeated antibiotic use in humans may also increase the reservoir of

antibiotic-resistance genes available to pathogens. For example, the microbiota of two healthy human adults harbored 115 unique inserts encoding transferable antibiotic-resistance genes, nearly half of which were 100% identical to resistance genes found in known pathogenic isolates [160]. In pigs, antibiotic treatment greatly increased the diversity of antibiotic-resistance genes over an already high background of resistance, even for classes of antibiotics that were not administered to these specific animals [161]. Similarly, when six human subjects were treated with clarithromycin and metronidazole (commonly used for treatment of Helicobacter pylori infections), the antibiotics greatly reduced gut bacterial diversity, and the communities remained perturbed four years after treatment in some individuals. Repeated and extensive antibiotic usage in humans thus likely selects an increasingly potent reservoir of antibiotic-resistance genes.

The impact of antibiotics, particularly during important developmental milestones, can be seen even when administered at subtherapuetic levels. In mice, subtherapeutic antibiotic treatment (STAT), commonly used to promote growth in domestic farm animals, led to increased adiposity and altered metabolic function [162, 163]. The combination treatment of penicillin and vancomycin, as well as treatment with chlortetracycline alone, significantly decreased the Bacteroidetes/Firmicutes ratio. This ratio has been previously associated with obesity and increased weight gain in wild-type mice [53], and in mice genetically predisposed to obesity [164]. The caloric output of fecal samples collected from STAT-treated mice decreased, consistent with the hypothesis that the gut

microbiota in STAT-treated mice extracts more energy from the diet than that in untreated mice [53]. The gene content of the microbiome was also affected: relative abundance of butyryl CoA transferase genes increased at 3 weeks, but recovered to baseline levels by 6 weeks. Relative abundance of formyltetrahydrofolate synthetase genes did not significantly differ at 3 weeks or 6 weeks, indicating that changes in gene levels are likely antibiotic-specific. STAT significantly upregulated genes involved in liver pathways associated with lipogenesis and triglyceride synthesis, perhaps leading to the observed increases in fat mass accumulation. This study is especially intriguing in the context of an epidemiological study of >11,000 children in the UK, which concluded that antibiotic use before 6 months of age was significantly associated with increased body mass between 10 and 38 months of age [165]. Thus, the developing microbiome of infants may be particularly susceptible to deleterious, long-lasting effects derived from antibiotic use.

**Cesarean Sections**

Maternal transmission has been shown to be a crucial factor in passing on protective microbes to offspring in many species. In Drosophila neotestacea, for example, the parasite Howardula aoronymphium causes near universal sterility in females and reduced mating success in males. In order to protect against this parasite, D. neotestacea transfer the bacterial endosymbiont Spiroplasma between mothers and eggs. In wild populations, females infected with Spiroplasma in addition to H. aoronymphium are more than ten times as fertile as H.

aoronymphium infected females that do not also harbor Spiroplasma [132]. In humans, the earliest exposure to foreign microbes for newborns has historically been from the vaginal microbial community during birth. This natural route of inoculation is bypassed in Cesarean sections, which are performed with increasing frequency worldwide despite evidence of significant deleterious effects [34, 166, 167]. A study of 165 Finish newborns (141 delivered vaginally, 24 delivered by Caesarean section) showed that by 1 month of age the C-section delivered infants had significantly less Bifidobacteria than did their vaginally-delivered counterparts, and also had significantly reduced bacterial cell counts in their stool [167]. Similarly, Swedish children who developed allergies by age 5 were less colonized by several Lactobacillus species (*L. rhamnosus, L. casei, L. paracasei*) and Bifidobacterium at birth [168]. Thus, vaginal delivery may inoculate a newborn with Lactobacillus and Bifidobacterium species that confer protective benefits later in life. Studies of exogenous inoculation of newborns with these important microbes in cases where C-sections are medically indicated are therefore needed.

**Exposure to Animals and Livestock**

Another mechanism by which modern humans may have lost some of their ancestral microbes is the reduced exchange between individuals and their environment, particularly through reduced exposure to animals. Ancient and rural societies typically have larger extended families that live with one another in close proximity; they also tend to have more contact with farm animals including

livestock, and with wild animals (e.g. those hunted as food), than do populations in more industrialized settings. In constrast, family units in many 'western' countries consist of only parents and their offspring living in one residence. These smaller households and decreased exposure to animals (other than domestic pets) likely reduce microbial transmission, including possibly beneficial microbes. For example, individuals living within a household share a greater proportion of their skin microbiota than non co-housed individuals [169]. Furthermore, the presence of a dog in the family facilitated the spread of rare, low-abundance microbes, including the family Methylophilaceae (class Betaproteobacteria) derived from canine oral communities, and families from within the Actinobacteria and Acidobacteria, likely derived from soil. The likely route of transfer was oral-skin transmission from the dog to household members. Exposure to animals, especially during the post-natal period, is especially important. In a study of 1,187 infants, Havstad et al. [136] found that IgE levels, typically elevated in diseases with an allergic component, were significantly lower in children who were exposed to pets early in childhood. These findings are consistent with the hygiene hypothesis, which states that exposure to certain microbes, including microbes obtained by pet-human transmission, trains the immune system to recognize foreign microbes and avoid harmless allergens.

**Replenishing the Host's Beneficial Microbiota**

Because deviations from a "normal" healthy microbiota are linked to many

human diseases, it is increasingly important to discover how to "reset" and "replenish" our gut microbiota with beneficial microbes (**Figure 3.2**). Different nutrients from the host's diet probably help determine which niches are available for microbial utilization, and thus which microbes become established. Large-scale changes, including a steady increase in microbial diversity, are seen in an infants' gut microbiome over the first few years of life, in part as a result of changes in their diet [158, 170]. The intrapersonal variation of the adult gut community is relatively stable over time compared to differences between individuals [38, 171], and a core functional profile of the microbiome is present even though the species that contribute the functions to this profile vary among individuals [17]. In mice, large changes in the gut communities result from dietary changes over the course of 1-4 days, though the effects are easily reversible [172]. However, in humans the timescale appears to be much slower, and long-term diet as measured by food frequency questionnaires over the course of a year, but not short-term diet experimentally manipulated over 10 days in a laboratory setting, seems to have a major effect [173]. Dietary alterations may thus play a role in achieving stable, long-term microbiome manipulation, as has been discussed elsewhere in detail [174].

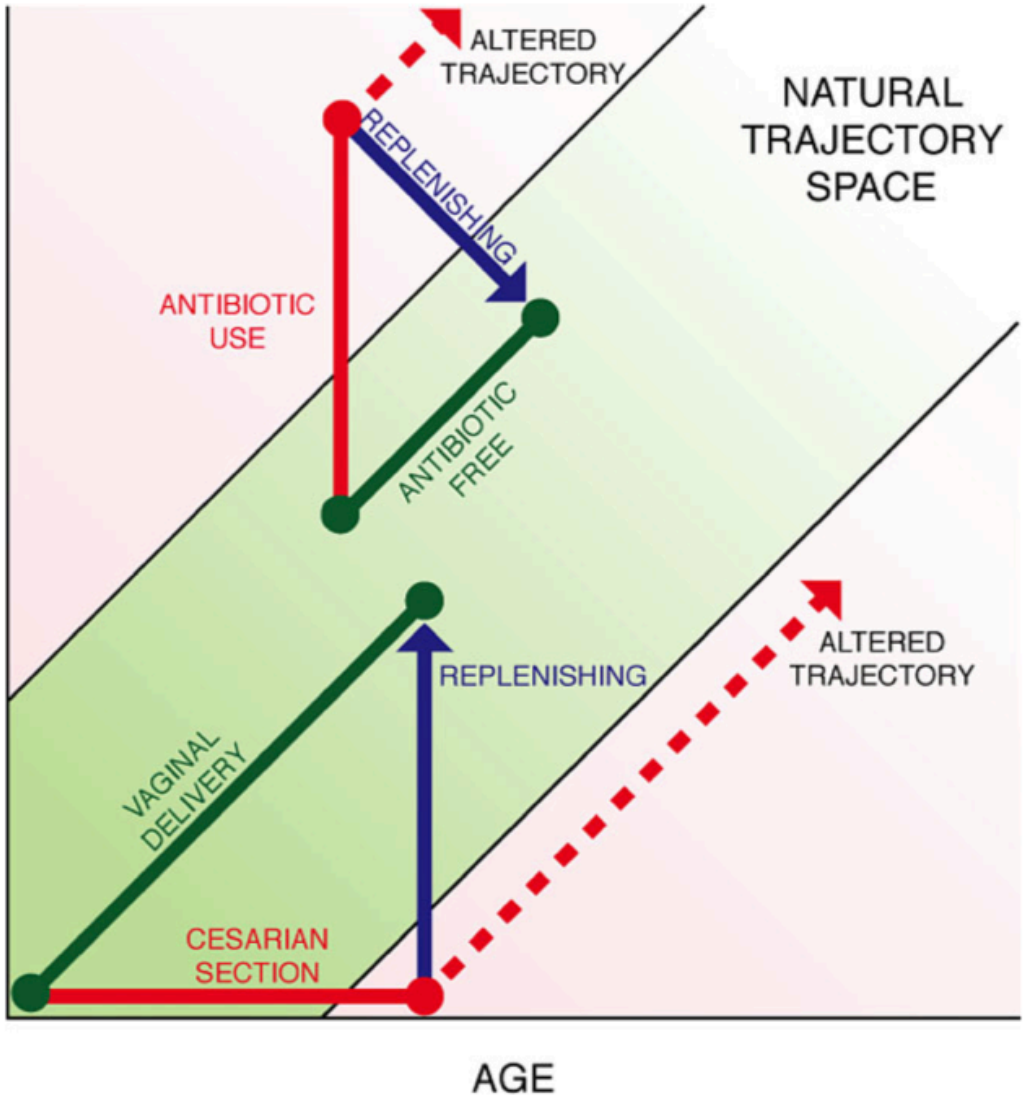


**Figure 3.2. Diverting our microbial communities back towards ancestral states.** Modern behaviors such as Cesarean sections and antibiotics may have the ability to push our microbial communities away from their natural, ancestral trajectories. However, microbiome manipulation may allow us to push our microbial communities back on track by replenishing the microbes that were affected by the disturbances. In this way, the impact on our microbial communities through events such as antibiotic use can be repaired such that our microbial communities maintain their protective benefits.

One of the best experimental systems for identifying members of the microbiota that are causally responsible for change is the method of personalized culture collections transferred into gnotobiotic mice [134, 175]. An individual's stool sample can be serially diluted, cultured from single progenitor cells without interference from other, faster-growing microbes, the individual strains can be characterized, and communities mirroring the original community can be reassembled. By reintroducing specific sets of taxa back into germ-free mice, the effect of the gut microbiota on host physiology can be determined directly, including the possibility of adding or removing specific members thought to be important. This technique will allow researchers to discover which taxa, or consortia of taxa, are required for preclinical efficacy in mice, and will guide clinical trials.

Because antibiotics profoundly reduce gut microbial diversity, it seems reasonable that antibiotic pretreatment might assist establishment of a new microbial community. Counterintuitively, antibiotics may actually impede the establishment of new communities. For example, antibiotic pre-treatment impaired the establishment of many phylotypes in rats after cecal transplant. Only 12 phylotypes of the input community were readily established across all antibiotic-pretreated rats, whereas 22 phylotypes were reproducibly established in the transplantation-only recipient rats (without antibiotics) [67]. The finding that cecal transplantation increased the overall diversity of non-antibiotic treated rats, and that this diversity was maintained beyond three months, shows that the gut may be more amenable to manipulation of the microbiota than previously thought. Another

example of microbiota remodeling comes from the observation that mice with reduced bacterial diversity after cefoperazone treatment recovered their full diversity when caged with normal mice, presumably assisted by coprophagy. Stool transplantations may thus help a gut community recover effectively even after antibiotic treatment [176]**.** These conflicting reports might be the result of individualized responses of community alterations following antibiotic treatment, as has been demonstrated in humans [61].

Probiotics – live microbes that, when ingested, have health-promoting effects – have also been used to treat individuals with gastrointestinal diseases (see Table 1 in [177]. However, public enthusiasm for probiotics has greatly outpaced the evidence of efficacy, and the hypothesis that probiotics affect the structure of the gut microbial community is not well supported by existing data. Consistent oral intake of the commonly-used probiotic strains *Lactobacillus delbrueckii spp bulgaricus, Lactococcus lactis spp cremoris, Bifidobacterium animalis spp lactis,* and *Streptococcus thermophilus* in humans did not significantly alter the gut microbiome in terms of community composition, structure, or gene content [178]. However, the probiotics did up-regulate bacterially encoded pathways involved in polysaccharide degradation in fecal and urinary samples [178]. Thus, probiotics might convey health benefits in some cases by modifying gene expression in the host and/or microbiota, rather than by changing the composition of the microbiota itself. Given the importance to infants of developing a healthy microbiota, it has been suggested that probiotics could place the infant's microbiota back on track

developmentally when altered by antibiotics early in life (**Figure 3.2**). However, these first few years of life include crucially important developmental processes, as also demonstrated in mouse studies showing that early interaction with the microbiota can permanently affect brain development and behavior [179]. Therefore, there is substantial risk of unintended consequences and caution should be exercised [180]. Future research should seek to understand why and how our gut microbiome changes, understand the functional consequences of those changes, and develop new therapies to return our microbiome to a healthy state.

**Fecal Transplants**

The evidence that out gut microbiota is important for educating our immune system is compelling, and modern behaviors may limit our exposure to specific and important microbial "teachers". Can we replenish our microbiota to compensate for this loss? The best case study for beneficial manipulation of the microbiota can be seen in the increasingly popular use of fecal microbiome transplantations for the treatment of recurrent Clostridium difficile infection. C. difficile infections, inflammatory bowel disease, and irritable bowel syndrome have all been associated with dysbiosis of the host's gut microbiota, leading to recurrent inflammation, diarrhea and constipation, although the mechanism of pathogenesis remains unknown [181, 182]. C. difficile is the most common cause of diarrhea associated with the use of antibiotics; the antibiotics permit particular C. difficile strains to dominate the community and release toxins A and B, which promote diarrhea [183,

184]. Recurrences in diarrheal episodes are generally treated with antibiotics; however, up to 65% of patients receiving antibiotics suffer relapse [185].

In contrast to the general ineffectiveness of antibiotics for treatment of C. difficile infections, fecal transplantation is highly effective both in animal models and in humans. For example, Lawley et al. [134] infected mice with C. difficile, resulting in a chronic intestinal disease. When treated with vancomycin alone, the C. difficile returned within 5-7 days of antibiotic cessation. Conversely, when the infected mice received a fecal transplantation from a healthy donor, the C. difficile infection did not return even months after treatment in 23 of 25 mice. To test whether the whole community was required or whether a lower-diversity subset would be sufficient for recovery, the authors cultured a healthy microbiota fecal sample through several generations (or passages), to reduce the community to only its culturable members. C. difficile infected mice were successfully treated using communities that underwent Passage 1 and 2, which already were reduced in phylogenetic diversity, but not Passage 3, where the community was very low diversity and dominated by Enterococcus spp. and Enterobacteriaceae spp. These experiments showed that the full community diversity of a gut microbiota is not required for clearing persistent C. difficile infections, but that replenishing the gut with specific members of the microbiota drive the transition from a diseased to healthy state. Ultimately, the authors identified a minimal mixture of 6 phylogenetically diverse taxa consisting of three novel species of Bacteroidetes sp. nov., Enterorhabdus sp nov., Anaerostipes sp. nov., and the previously identified

Lactobacillus reuteri, Enterococcus hirae, and Staphylococcus warneri, that could resolve C. difficile infections. This study underscores the current interest in the intersection between personalized medicine and microbial ecology for identifying communities that can modulate health status: simple communities with culturable members provide the advantage that they can be more easily packaged, characterized and dispensed, but the full diversity of the community may be required for some disorders.

## Conclusion

The field of human microbial ecology is evolving, and has recently transitioned from demonstrating that specific microbial consortia are associated with disease states towards learning how to directly manipulate the human microbiome for therapeutic purposes. The use of whole fecal transplants and highly defined microbiota transplants for C. difficile infections has demonstrated that microbiome manipulations can achieve high efficacy in at least one case where traditional pharmaceuticals fail. Furthermore, they suggest that mouse models represent a highly tractable system for investigating microbiome manipulation that can then guide clinical applications in humans. In the future, antibiotics might be used to treat the most severe infections, but their long-term effects on the microbiota may be mitigated by reintroducing species from the same person in a state of health, from other people (and perhaps from populations living more traditional lifestyles), or from engineered microbial consortia. Replenishing our

defensive microbes will allow us to direct our microbiota back onto its evolutionary tracks, which may be especially important given that modern behaviors and practices likely create microbial detours not previously encountered in our evolutionary history.

# Chapter 4

## Applied Microbiome Analysis of Hunter-Gatherers

This section includes my contributions to a microbiome study on a hunter-gather society called the Matses who live in the Peruvian Amazonian jungle. Sample collection and preparation was conducted by members of Cecil Lewis's lab at the University of Oklahoma. The figures and results included in this section are the direct result of the statistical and microbial analysis that I performed and wrote as a part of the final submitted manuscript.

### Introduction

Even with the advent of cost-effective sequencing methods, a full characterization of the human gut microbiome across evolutionary history remains incomplete. People living in remote communities, especially hunter-gatherer societies, are the best modern-day sources to examine the diversity and structure of the gut microbiome of ancient humans. An understanding of our ancient human gut is important for assessing how modern day interventions and perturbations are possibly pushing our gut microbiomes off of their natural, evolutionary trajectories. This work represents the microbial community and functional gene profile characterization of the Matses, a group of hunter-gatherers living in the Peruvian Amazonian jungle, who are the best living source of humans living a lifestyle of our ancient ancestors.

**Matses Harbor a Unique Gut Microbiome**

We compared the overall microbial community structures using unweighted UniFrac and principal coordinates analysis (**Figure 4.1A**). The non-western populations clustered distinctly in PCoA space compared to the US population, and within the non-western communities, the Matses were distinct ($P = 0.001$ and $P = 0.001$, respectively, PERMANOVA). Matses individuals resembled each other to a greater degree than did individuals from the US or Venezuela, as UniFrac distances for the Matses were much smaller than the US or Venezuela (**Figure 4.1A**). The machine learning method of random Forests, utilizing microbial community signatures, was able to correctly identify a Matses individual from someone from the United States, Venezuela, or Malawi, with only 1.9% error. We were unable to find significant taxonomic differences in the Matses population between individuals affected or unaffected by parasites, suggesting that immune modulation by parasites does not detectably affect the microbial community in this setting.

**Matses Populations are Lower in Microbial Diversity Than Other Populations**

Previous reports have indicated that Western populations are lower in microbial diversity than non-Western populations. We therefore expected that the Matses would be especially diverse. Interestingly, this appears not to be the case: the Matses instead have different kinds of taxa, rather than more taxa,

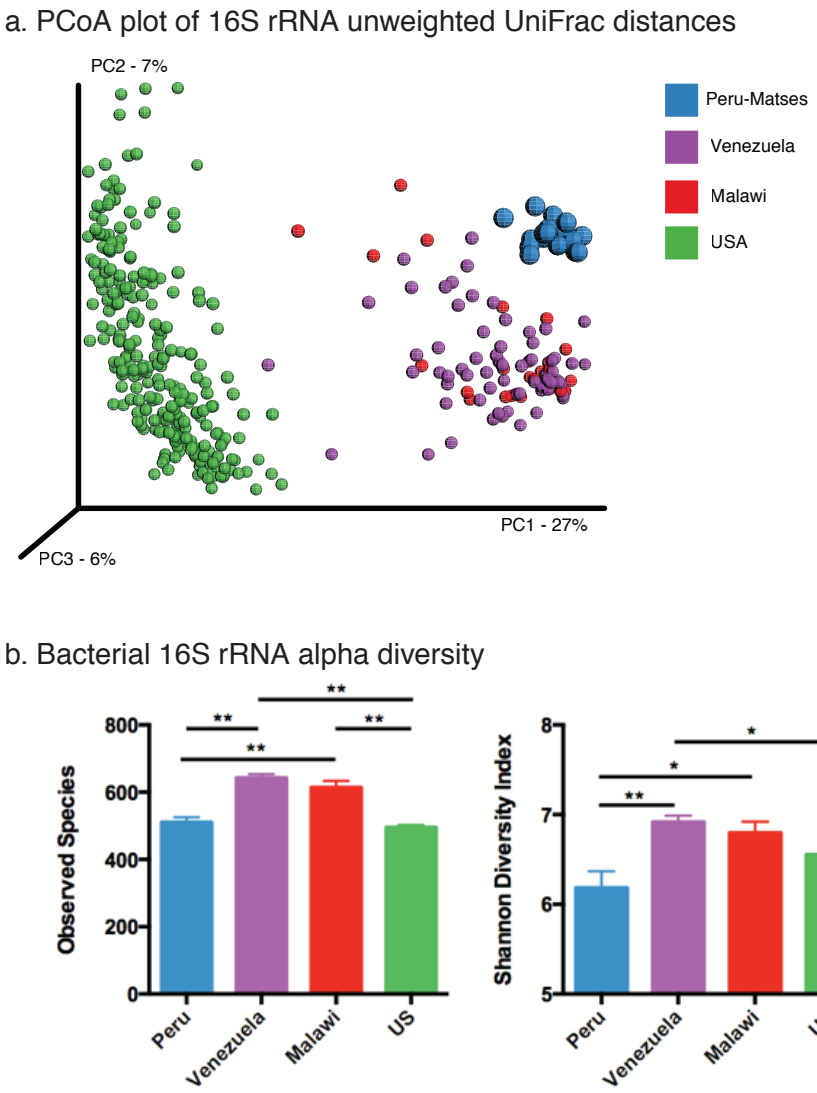


**Figure 4.1. The Matses clearly separate from the US, Venezuela, and Malawi populations.** A) This analysis shows principal coordinates analysis of unweighted UniFrac distances, taking into account the full phylogenetic tree. PC1, which explains the largest fraction of overall variance among the four populations, clearly separates out the Western communities (US) from the non-Western Communities (Peru-Matses, Venezuela, and Malawi). PC2, explaining the second largest portion of the variance, divides the Matses from Venezuela and Malawi in the non-Western portion of the PCoA plot. Venezuela and Malawi are separable along additional dimensions (data not shown). B) Bacterial alpha diversity across populations. A) Observed species B) Shannon diversity index. The Matses have the same alpha diversity as the US, while Matses have significantly lower alpha diversity then both their non-Western counterparts of Malawi and Venezuela. Bars indicate mean +- s.e.m. *P < 0.05, **P < 0.01 (nonparametric t-test with 1000 Monte Carlo comparisons, Bonferroni corrected). We show results of 16S rRNA V4 region data generated on the Illumina Hiseq 2000 platform.

than other populations studied to date. To eliminate the possibility that any differences in diversity were due to the Matses harboring OTUs absent from the reference database (which is biased towards US samples), approximately 1 million sequences from each of the four communities were randomly subsampled from the total sequence pool and then clustered with an open-reference procedure that includes novel sequences in the analysis. Children below 3 years old were excluded to avoid the different infant microbiome as a confounding factor. Surprisingly, the Matses had significantly less diversity both in the number of different OTUs and the Shannon Diversity Index compared to the other rural populations (**Figure 4.2B**), and, by the Shannon measure (which is influenced by the evenness of the distribution), are lower in diversity even than the US population.

**Functional Characterization of the Matses Microbiome**

Given that taxonomic differences have been correlated with functional differences in a range of populations [186, 187] we collected shotgun metagenomic data from the Matses to assess the functional capacity of this unique and isolated hunter-gatherer community versus the agrarian cultures of Venezuela and Malawi and the Western US. We combined our metagenomic data from 21 Matses individuals with reads obtained from Yatsunenko et al. (2012), representing shotgun reads from Malawi, Venezuela, and the US, and re-annotated the entire metagenomic dataset together for consistency in gene identification. The KEGG Orthology (KO) profiles of the four communities are distinct when assessed with

principal coordinates analysis (**Figure 4.2,** P = 0.001, PERMANOVA), mirroring the findings seen in the 16S rRNA data. Procrustes analysis, which tests fit in the spatial orientation of data from two different principal coordinates plots, showed good agreement between the taxonomic and the functional datasets on 52 matched samples (**Figure 4.3A**), suggesting that, as in other populations, taxonomy is a good guide to function. We then compared the relative differences between the functional KO profiles and the 16S taxonomic profiles between countries (**Figure 4.3B, Figure 4.4**). When the full compliment of KOs and taxa are considered, the Matses are significantly different from the US both functionally and taxonomically, but not different from the non-Western populations.

To determine if the relative abundance of KOs involved in metabolism were significantly different between populations, we then filtered down the KOs to only those shared by every sample from all populations, and ran Kruskal-Wallis. It is clear that the Matses harbor increased metabolic capacity compared to the US and non-Western populations in these shared KOs, especially in nucleotide, amino acid, and energy metabolism. We discovered 84 of the 121 significantly different metabolism KOs between Matses and the US had higher relative abundance in the Matses community (**Table S4.1**). Similarly, Matses had a higher relative abundance in 78 out of 104 shared KOs associated with metabolism compared to the non-Western Malawi and Venezuela populations (**Table S4.2**). The Matses also display the increased glutamate synthase capacity seen in the non-Western of Venezuela and Malawi compared to the US

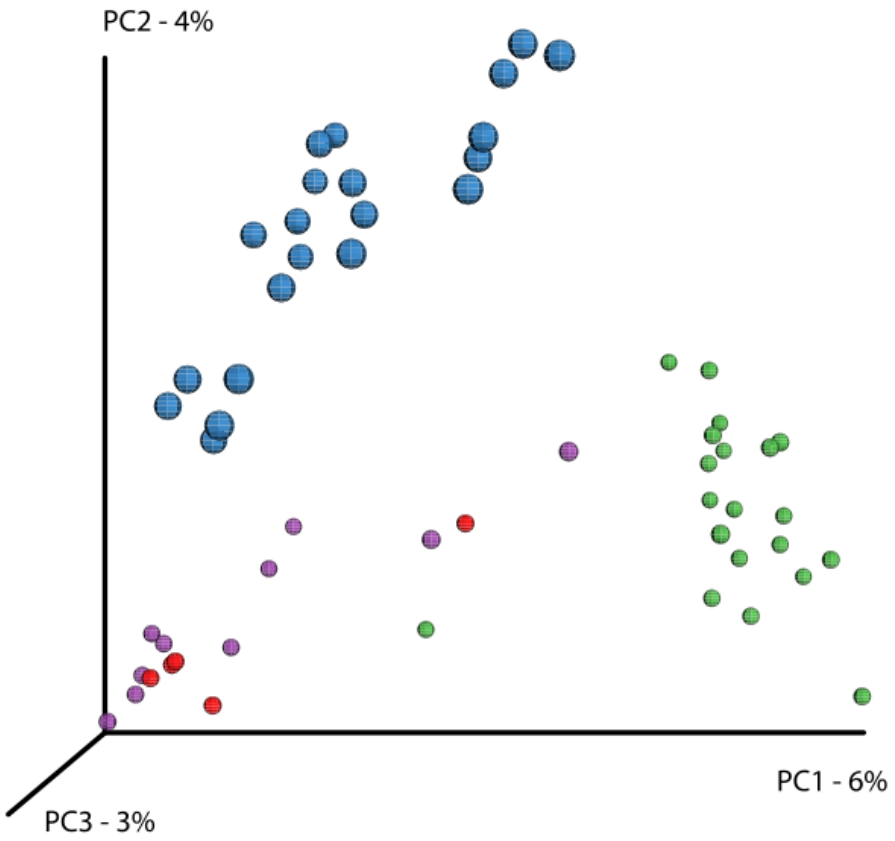


**Figure 4.2**. **PCoA plots showing Hellinger distances between all four populations for KOs.** Data derived from shotgun sequences that were filtered down to 50,000 counts/sample for adults (age > 3 years) only. Blue spheres represent Matses, purple spheres represent Venezuelans, red spheres represent Malawians, and green spheres represent individuals from the United States.

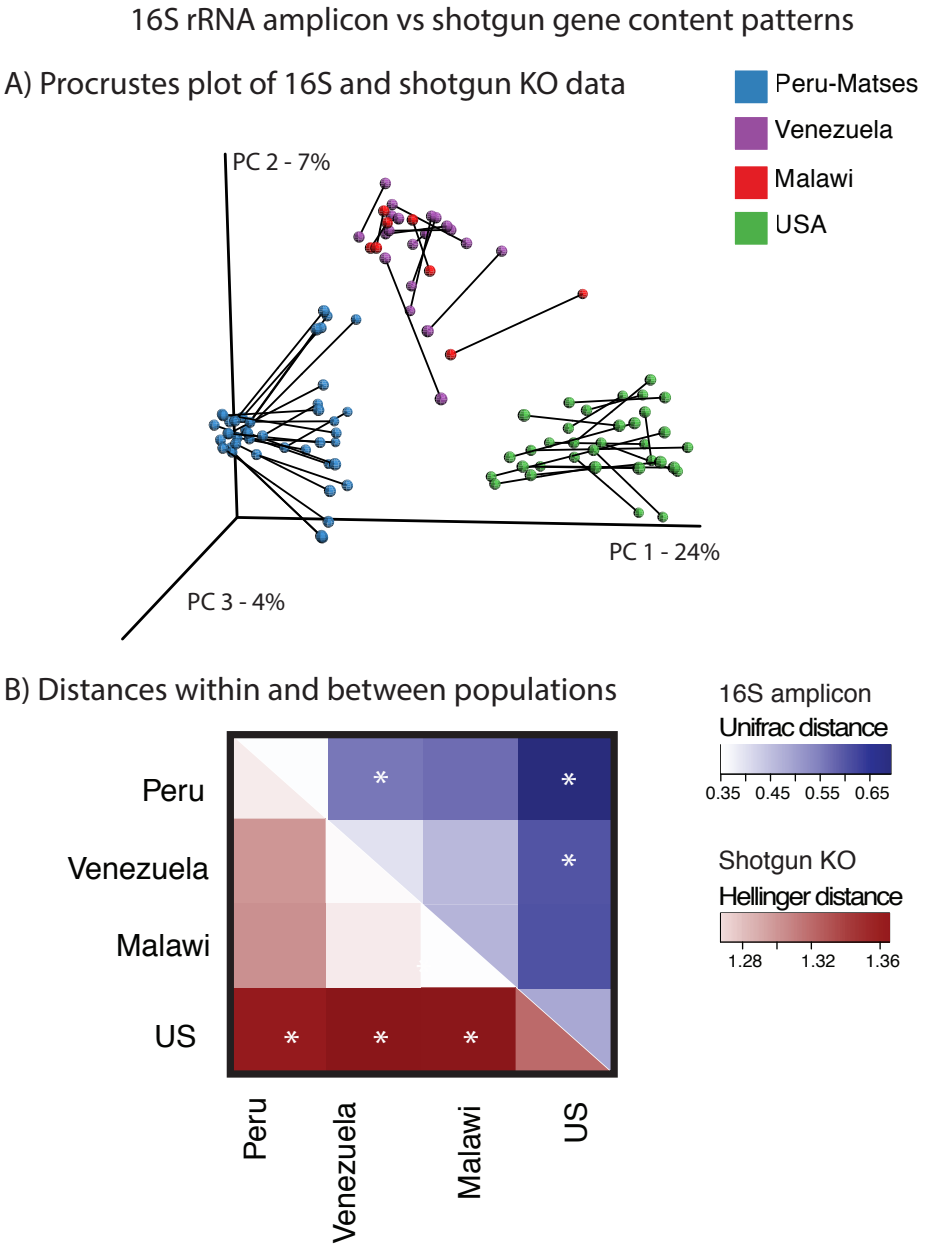


**Figure 4.3. 16S and KO profile comparison.** A) Between- and within-country differences assessed by unweighted UniFrac for 16S V4 rRNA data and Bray Curtis for shotgun metagenomic data. Only individuals older than three were included (n=21). The diagonal represents the within community distance. Significant differences (p-values<0.05) are denoted by an asterisk. B) Procrustes analysis between the taxonomic and the functional datasets on 52 matched samples

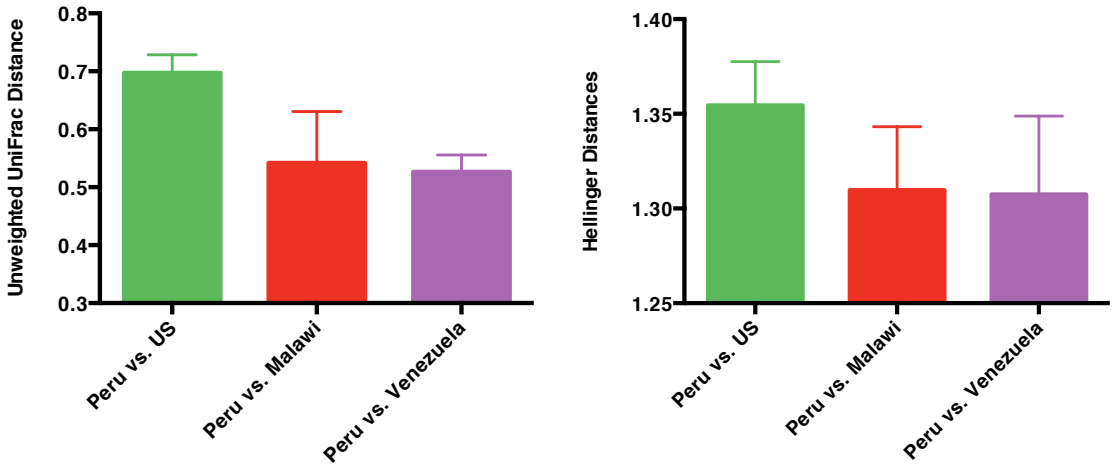


**Figure 4.4. Pairwise comparison of Matses 16S unweighted UniFrac distances and shotgun gene content Hellinger distances**. Matses are compared to USA, Malawi, and Venezuela. Error bars display standard deviations. Individuals from USA are most different from those in Matses in both 16S profiles and functional gene content. Venezuela and Malawi individuals had 16S profiles and functional profiles that were equally distant from Matses. We included data from 52 samples for which we had both 16S rRNA and shotgun metagenomic data.

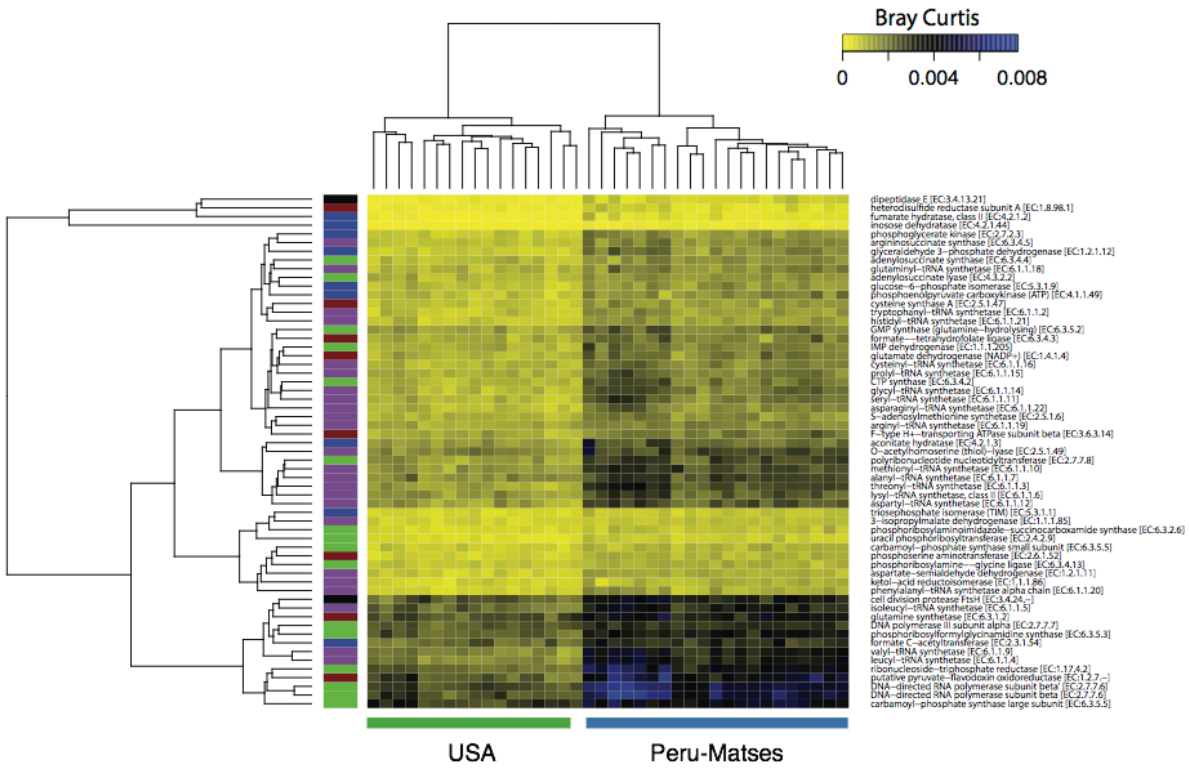


**Figure 4.5**. **Heatmap of KOs that are significantly enriched in Matses compared to US.** Sidebar is colored according to L2 KEGG Orthology: Amino Acid Metabolism = Purple, Carbohydrate Metabolism = Blue, Energy Metabolism = Red, Enzyme Families = Black, Nucleotide Metabolism = Green. Dendrograms represent average Bray Curtis distances for samples and KOs. The samples perfectly divide into Matses (on the right of the heatmap) and US samples (on the left).

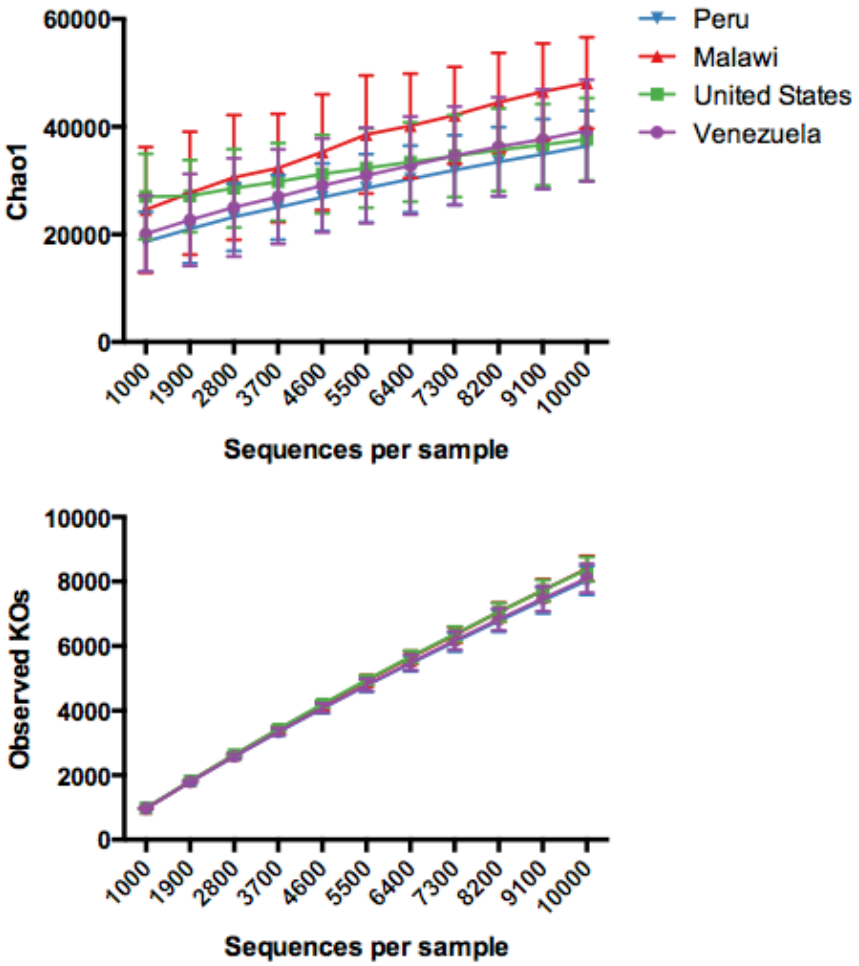


**Figure 4.6. Rarefaction plots of KOs.** Rarefaction plot (1,000 – 10,000 Kos per sample) of Chao1 and the number of observed Kos for each country.

(**Figure 4.5**) [187]. Considering the full KO data set, Matses did not have a significantly higher diversity of KOs according to the chao1 metric or by number of unique KOs (**Figure 4.6**). One possible explanation for this is that our shotgun metagenomic data was annotated against databases typically curated from Western samples, and therefore any KOs unique to the Matses are not likely to be uncovered until significant genome construction has been accomplished. It therefore seems that while the Matses may not have an increased number of unique KOs associated with metabolism, but a higher proportion of their microbiome is related to these functions.

## Conclusion

Understanding microbial communities and their functions in populations living relatively ancestral lifestyles is essential for understanding the coevolution of humans as a species with our microbiomes. As seen in Chapter 3, understanding of our ancestral humans might provide insights into how current societal disturbances are altering the natural evolutionary trajectory of our microbiomes, and possibly contributing to disease states. This work demonstrates that a group of hunter-gathers harbor a low-diversity gut microbiome, but without the resulting increases in obesity common in modern societies. Furthermore, this group of individuals has a microbiome that is enriched for metabolism genes, likely resulting from their need to pull maximum energy from their sparse and unpredictable food supplies.

# Chapter V

# Microbiome-Metabolomics Interactions

This published chapter introduces the interactions that take place in the gut between the microbiome and the metabolome, including implications for host health [188].

## Introduction

Rapid advances in sequencing technologies over the past decade have allowed researchers worldwide to assess how the intestinal microbiome affects human health[189]. Humans develop symbiotic relationships with microbes at a young age[33]. Factors such as the environment[128], proximity to other humans and animals[169], diet[173, 190], genetics[191], and temporal variation[38] affect the assemblage of microbes on our skin, in our mouths, and in our guts[130, 171]. Our microbiota has been compared to a previously unknown organ in terms of its effects; it has extensive metabolic capabilities, and carries ~150-fold more genes than the human genome. Microbes provide the host with a range of otherwise inaccessible metabolic capabilities[5].

Unlike the human genome, the microbiome is relatively plastic. It can be rapidly altered through factors such as diet[190], drugs, probiotics, and microbially produced metabolites. Deliberate alterations in the microbiota and/or microbiome can therefore affect health. The intestinal microbiota is viewed increasingly as an

important target of pharmacologic agents—specific microbes have been shown to deactivate or activate specific xenobiotics, which can alter the effects of different therapeutic agents[192]. The systems-level effects of the entire microbial community on the whole metabolite repertoire are just beginning to be understood.

Metabolomics and metabolite profiling analyses have been widely used to identify disease biomarkers. For example, quantification of triglycerides, glucose, and cholesterol in the blood can be used to determine the risk of heart disease. Similarly, the first microbiome studies sought to identify taxa that correlated with disease, physiological state, drug use, or dietary intake. However, not all exposures can alter the composition of the microbial community or its gene content; some can affect gene expression[178, 193].

Humanized mice (created by transplanting human fecal microbiota into the mouse gut) have metabolomes distinct from those of conventionally raised mice[194]. This observation indicates that different gut microbes can produce changes in metabolites throughout their host. This shift in focus from determining "who is there" toward understanding "what are they doing" drives current studies of the human microbiota. Metabolomic studies will allow us to move from observing patterns to understanding mechanisms.

Metabolomic analyses also help researchers to understand the effects of rare taxa, and taxa with genomic variations that affect function. Organisms are considered to be of the same species if they have greater than 97% identity in the 16S rRNA gene. However, genomes from the same species can have large

differences in DNA sequences outside the 16S rRNA gene. Importantly, they often have different sets of gene clusters that regulate production of specialized metabolites (e.g. antibiotics, virulence factors, siderophores, etc.) and the composition of the microbial communities, as well as encode many antibiotic resistance genes[195]. Rasko et al. determined that among 17 *Escherichia coli* isolates, the average genome size of a single isolate was 5020 nucleotides (nt), although the pan-genome was ~13,000 nt[196]. Furthermore, rare taxa might have a large effect on the overall community metabolome if they have important metabolic activities, perhaps acting as keystone species.

Although definitions of what constitutes a core microbiome in terms of membership is elusive, there does seem to be at least a core functional profile for the gut microbiota.[130] Identifying biologically important variations against this core remains a challenge. Metabolomic analyses provide a partial picture of metabolism rather than the potential for metabolism, and the expression of this core set of functions can change with alterations in available substrates, such as xenobiotics, even if the microbial species membership and abundance remain constant[193]. We review the intimate connections among animal hosts, their microbiota, and the metabolites produced by either one.

Different microbial communities metabolize xenobiotic agents and dietary components in different ways to produce variable effects on many tissues in the host, including the brain[197] (**Figure 5.1**).

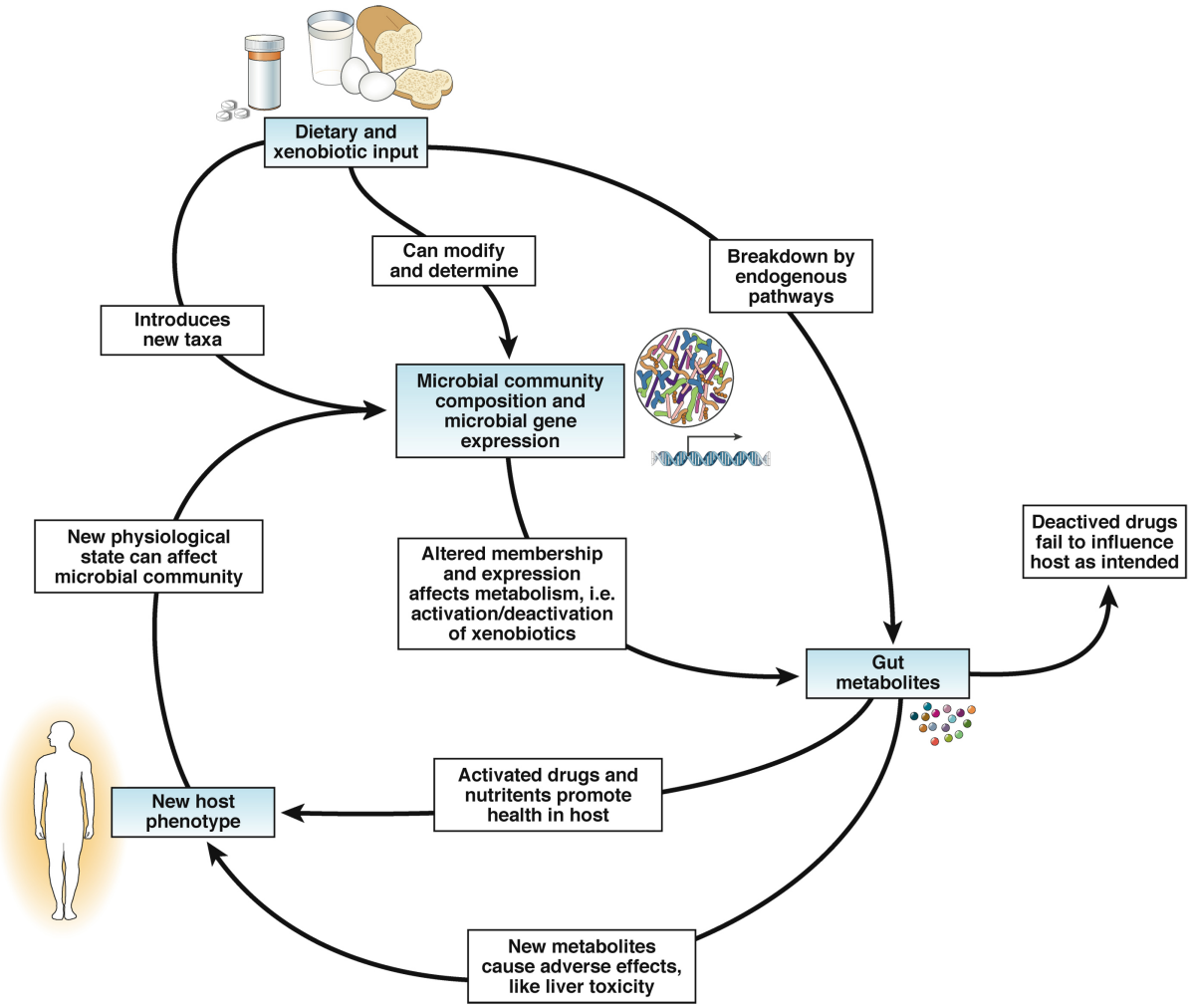


**Figure 5.1**: **Interactions among host, microbiota, and metabolites**. In this simplified model, the gut microbiota metabolize substrate inputs from the host including diet and xenobiotics into metabolites that can enter the host's bloodstream and affect the host peripherally. For example, therapeutic drugs can be inactivated, reducing their efficacy. Alternatively, drugs may converted to derivatives with non-target and possibly toxic effects. Changes in these input substrates, therefore, change the reservoir of available microbial substrates and alter the metabolomic profile of the gut, yielding variable effects on the host. The new host phenotype can, in turn, have a feedback effect on the microbial community.

We discuss general metabolomic technologies and their implementation for study of human health, assess cases in which changes in gut microbiota alter host metabolic profiles, examine the ways in which gut microbiota process xenobiotics and nutritional inputs, and examine the analytical limitations of associating microbial abundances with metabolic profiles.

**Metabolomics in Assessment of Metabolic Status**

Metabolomic studies analyze complex systems, including the repertoire of small-molecule metabolites in the gut, using high-throughput analytical methods. Mass spectrometry and nuclear magnetic resonance spectroscopy allow robust and sensitive identification of metabolites produced by microbes and host cells, in samples such as feces, urine, and tissue (see comprehensive reviews in [198, 199]). These tools allow researchers to determine the effects that treatments or perturbations have on the host's metabolic profile, by analyzing the presence and quantity of thousands of metabolites simultaneously. Although it is a challenge to assign spectral features, spectral networking platforms,[200, 201] aided by open-source metabolome databases such as HMDB[202], METLIN[203], LIPIDS MAPS[204], MassBank[205], and NIST,[206] allow for faster identification and annotation of known and unknown metabolites[207]. By comparing pre- and post-perturbation metabolomic profiles using multivariate statistics, metabolites that are significantly affected by experimental variables can be identified and placed into the larger context of how the host was affected overall.

**Effects of the Microbiome on the Metabolome**

Metabolomic analyses allow for the metabolism of the gut microbiota to be directly compared with metabolic outcomes in the host. Wikoff et al.[208] directly tested the effect of gut microbiota on the host by comparing the plasma metabolomic profile, obtained via untargeted mass spectrometry, between germ-free and conventionally raised mice. They found that concentrations of more than 10% of all metabolites detected in the plasma differed by at least 50% between mice with and without gut microbes. Furthermore, many metabolites were detected only in serum from conventionally raised mice (not germ-free mice). For example, serum levels of tryptophan decreased 40% in serum from conventional mice compared to germ-free mice—likely due to the presence of bacteria that produce tryptophanases[208].

Another detailed study evaluated the systemic effects of probiotics, prebiotics, and their combination (termed 'synbiotics') in initially germ-free mice colonized with a combination of microbes representing those found in a human infant (*Bacteroides distasonis*, *Clostridium perfringens*, *Escherichia coli*, *Bifidobacterium breve*, *Bifidobacterium longum*, *Staphylococcus aureus*, and *Staphylococcus epidermidis*)*[209]*. Dietary supplementation with the probiotic *Lactobacillus rhamnosus NCC4007* and the prebiotic galactosyl-oligosaccharides significantly altered the relative proportions of the 7-member community, and led to systemic changes in the metabolic profiles of different tissues from the mice. For example, a prebiotic increased proportions of *B breve*, *B longum*, and *B distasonis*; decreased

proportions of *E coli* and *C perfringens*; and altered lipid metabolism by reducing plasma levels of glucose and hepatic levels of triglycerides. Probiotics also had systemic effects, lowering plasma levels of lipoprotein, hepatic levels of glutamine, and glycogen levels. Overall, prebiotics significantly altered the metabolome in the plasma, urine, feces, liver, pancreas, renal cortex, renal medulla, and adrenal glands; probiotics produced differences in all these compartments except the pancreas.

Interestingly, another study that evaluated the effects of probiotics and prebiotics in adults found that neither significantly affected proportions of microbes in fecal samples, but RNA sequencing data showed altered expression of microbial genes that control carbohydrate metabolism[178]. It is possible that the relatively simpler communities that reside in infants are more susceptible to probiotic and prebiotic manipulation than the more diverse and complex communities found in adults. Prebiotics and probiotics might therefore have the largest effects when administered early in life. However, this hypothesis requires testing in animal models.

The dietary components that escape digestion in the upper gastrointestinal tract provide most of the substrates for the intestinal microbiota. Fermentation of carbohydrates by the intestinal microbiota leads to the production of short-chain fatty acids (SCFA) such as butyrate, propionate, and acetate. Studies have shown that patients with inflammatory bowel diseases such as ulcerative colitis have fewer butyrate producing bacteria (e.g., *Roseburia hominis* and *Faecalibacterium*

*prausnitzii*) in their intestine, resulting in lower levels of butyrate[210, 211]. In addition to butyrate, propionate can potentiate de novo generation of T-regulatory cells in the peripheral immune system. Modulation of butyrate- and propionate-producing microbes might therefore be used to treat inflammatory bowel diseases such as ulcerative colitis. However, the anti-inflammatory mechanisms of butyrate and other SCFA remain poorly defined.

**Predictive Microbial Metagenomes**

Metagenomic information can been used to determine how metabolism is affected by different disease states. Studies of obesity have shown that individuals with increased adiposity have lower microbial diversity than lean individuals [17, 212]. The more-diverse microbiota of lean individuals contains significantly higher proportions of microbes correlated with anti-inflammatory responses, such as *Faecalibacterium prausnitzii*. The less-diverse microbiota of obese individuals contains higher proportions of *Bacteroides sp.* and *Ruminococcus gnavus*, each of which could have inflammatory effects[212]. Gene content analysis of these groups revealed the less-diverse microbiota appeared to produce lower levels of butyrate, have increased potential for production of hydrogen sulfide, and have reduced capability for management of oxidative stress. One poorly understood aspect of the microbiome, and its potential to produce a variety of metabolites, is whether microbial diversity itself has protective effects for the host, or whether low diversity

is a side effect of specific disorders (rather than a cause)[213]. This relationship can best be resolved in humans by prospective longitudinal studies.

Although it would be ideal to obtain metabolomic and metagenomic data for every sample for which a 16S amplicon profile has been collected, these techniques are currently far more expensive than 16S amplicon profiling. Fully matched datasets are therefore prohibitively expensive and time consuming to produce. However, recent advances in software, including Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt)[214], that exploit the strong association between phylogeny and function now allow researchers to estimate the metabolomic functional profile of a community using 16S amplicon sequences. Briefly, PICRUST takes a phylogenetic tree where the gene profile of a subset of nodes is known, and then uses ancestral-state reconstruction to estimate the functional gene content for other uncharacterized nodes. PICRUSt was able to make strong predictions (average Spearman $r = 0.82$) for inferred metagenomes from 16S marker genes, compared against fully sequenced metagenomes obtained from the Human Microbiome Project.

Another powerful computational tool is Predicted Relative Metabolic Turnover (PRMT), which uses gene number to predict the relative consumption and production of metabolites in a system; it can be used for modeling and hypothesis generation[215]. Tools such as PICRUSt and PRMT could be cost-effective methods to determine whether additional resources should be used for more comprehensive

metabolic profiling and metagenomic sequencing. However, findings must be validated with matched datasets, to assess the limits of their performance.

**Metabolomic Profiles of Infants**

Changes to the microbiome and immune system during infancy may have lasting effects, such as in contributing to the development of allergies[178, 216, 217]. Distinct changes in the microbiota occur during the first 2 years of life, and correlate with changes in environment, diet—these can be tracked by studying changes in infants' fecal metabolomes. A study that followed infants at risk for celiac disease showed that the metabolomes of infants less than 6 months old were dominated by sugars, including lactose and glucose. However, after 6 months, their metabolomes shifted, increasing concentrations of amino acids and SCFA. Principal coordinates analysis showed that the metabolome of infants at 2 years of age resembles more closely that of adults, due to increased levels of acetate and butyrate[218]. These findings are supported by 16S amplicon studies showing that the infant microbiota comes to resemble that of adults from the same community at 2 years of age [157]. It is also apparent that the intestinal microbiota of infants is specifically adapted to metabolize the infant's earliest nutrient source, breast milk. Specific *Bifidobacterium* species have genomes enriched in genes that regulate processing of human milk-derived oligosaccharides. These might have a competitive advantage that places them among the first colonizers of the human intestine[219].

**Xenobiotic Metabolism**

In addition to diet-derived macronutrients, the microbes residing in the gastrointestinal tract may be exposed to a variety of xenobiotic compounds (antibiotics, other drugs, and diet-derived bioactive compounds). Because the gut microbiome encodes so many enzymes with different activities, it is not surprising that many of xenobiotic compounds are often metabolized by the gut microbiota. It has been at least 40 years since we began to appreciate the contribution of microbes to xenobiotic metabolism[220-222]. However, we are only beginning to uncover the mechanisms of this process. Adding to the complexity of these interactions, xenobiotics can also modulate the expression and activity of the gut microbiome[193]. Metabolites of microbial origin may interfere with host metabolism of xenobiotics, and diet-derived nutrients can regulate microbial metabolism of xenobiotics.

One of the first studies to provide detailed evidence for the interaction between the gut microbiota and metabolism of xenobiotics came from Clayton et al. in 2009[223]. Their study leveraged a powerful metabolomic analysis pipeline to correlate the presence of the microbial metabolite, *p*-cresol, with a reduction in the ratio of sulfonated to glucuronidated acetaminophen. Increased *p*-cresol production reduces the capacity of the liver to properly metabolize this widely used analgesic drug, presumably because *p*-cresol competes with sulfotransferase [223]. Subsequent studies from this group showed that metabolites of microbial origin could modulate expression and activities of a range of host enzymes, including those

of major xenobiotic-metabolizing cytochrome enzymes[224]. These seminal observations are beginning to lay the foundation for a metagenomic approach to selection of therapy based on microbial and host metabolism.

In addition to its interactions with metabolite production, the gut microbiota can also have a more-direct role in xenobiotic metabolism, by catalyzing a multitude of reactions that influence the fate of these compounds. Recent reviews have summarized the many processes by which microbes metabolize xenobiotics[192, 225, 226]. Although these activities are largely catalogued, there are only a few for which the exact mechanisms are being characterized. For example, it has been known for decades that the cardiac drug digoxin can be inactivated by *Eggerthella lenta*, a common gut bacterium within the Actinobacteria[227]. Researchers have recently identified a cytochrome-encoding operon that is upregulated by digoxin and other cardiac glycosides and is unique to strains capable of inactivating digoxin. Inactivation of digoxin was blocked by increasing dietary protein intake by mice mono-associated with *E lenta[228]*, likely due to the inhibitory effect of arginine[229].

Wallace et al. studied how the microbiota can determine the effects of the colorectal cancer drug irinotecan. Enzymes produced by microbes have long been known to deconjugate an irinotecan metabolite in the gut, causing inflammation, diarrhea, and anorexia. After a successful screen for a small-molecule inhibitor of the microbial b-glucuronidase enzyme that mediates this deconjugation, Wallace et al. showed that the side effects of irinotecan could be greatly reduced by co-

administration with this b glucuronidase inhibitor[230]. Interestingly, recent studies show that the presence of the microbiota increases the efficacy of chemotherapeutic drugs, indicating that the microbiota have previously unappreciated, but integral roles in mediating responses to these drugs [231, 232].

**Computational Challenges to Discovering Correlations**

Identifying statistically meaningful patterns in metabolite contingency tables (tables recording the abundance of each metabolite count in each sample) is straightforward in theory but often conducted with mathematically unfounded techniques in practice. For instance, analysis of variance and Student *t* test methods are frequently used to identify significant differences in abundances of metabolites among sample groups without establishing that the underlying data meet the distribution requirements. Normality, equality of variance, and homogenous population characteristics are required for proper calculations of statistical significance (either *P* values or false discovery rates). Although non-parametric tests can be substituted to deal with the non-normality of the data, these approaches still does not resolve 2 fundamental computational challenges: extraction of biologically significant results from the mass of statistically significant results and the fact that multivariate biological data are typically normalized to a sum—the simplex constraints this imposes violate the Euclidean-space models assumed by most test statistics.

The most widely applied method to reduce biologically irrelevant, but statistically significant, results is to remove features (taxa, KEGG Orthology groups, Enzyme Commission numbers, etc.) from the contingency table prior to testing on the basis of a metric that assigns expected biological relevance to a feature. This 'metric' is usually as simple as overall table abundance (e.g. remove feature *i* if *i* is less than 1% of all observations) or overall sample representation (e.g. remove feature *i* if *i* is in less than 20% of samples). This filtering approach is motivated by the intuition that more widely shared features will be more biologically important, and has the additional attraction of reducing the severity of multiple hypothesis test correction factors. Unfortunately, although widespread, this approach has not yet been systematically benchmarked or evaluated for sensitivity, specificity, or even false discovery control, particularly in fields combining microbiome and metabolomic datasets.

A complementary approach to identifying differential representation of features among groups is to look for interactions among features via co-occurrence analysis (**Figure 5.2**).

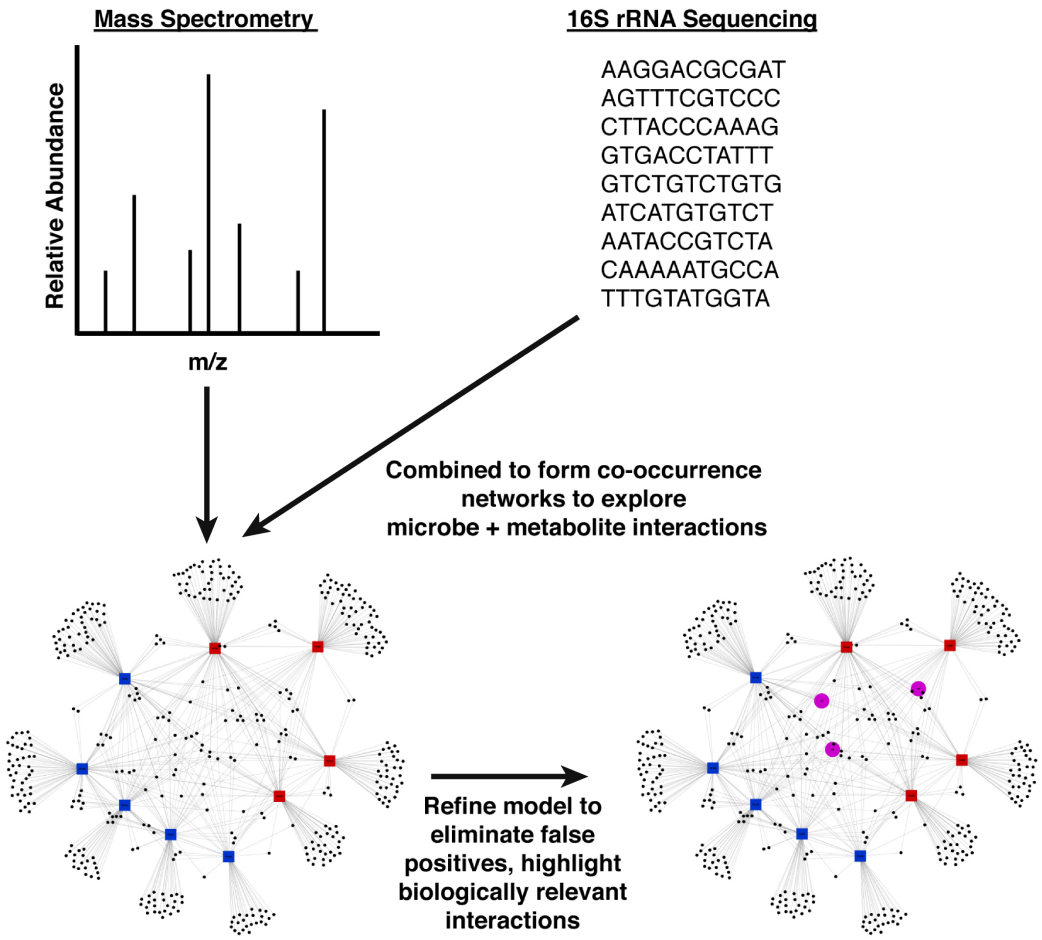


**Figure 5.2: Exploring the interactions between metabolomics and the microbiome.** Both metabolomics and high-throughput sequencing produce a wealth of information. Visualizing the interactions between these highly multivariate datasets is important for elucidating relationships. In this example tripartite network, the large blue nodes represent samples, which are connected to red diamonds (metabolites) with red edges, and connected to black circles (OTUs) with black lines. The closer an OTU node or a metabolite node is to a sample node, the larger the relative abundance of that metabolite or that OTU in that sample. Therefore, OTUs and metabolites that are close together in the network tend to be found in the same samples (and this suggests, but does not conclusively prove, that the metabolite may be produced by that OTU). The tripartite network also demonstrates which metabolites and OTUs are shared by samples, and which metabolites and OTUs are unique to a given sample. As discussed in this review, methods are being developed to help separate out biologically important associations from amongst many statistically significant ones. Once identified, we can visualize how biologically important metabolites are controlled by the interaction between host and microbiome.

Traditional co-occurrence detection methods including Spearman or Pearson correlation between feature vectors are not reliable when the data are 'compositional' (i.e. lie in a simplex rather than Euclidean space)[233, 234]. Because compositionality is a feature of much -omics data (16S amplicon surveys are inherently compositional because normalization for unequal sampling effort in any contingency table introduces compositionality), methods such as 'SparCC' and 'CoNet' have been developed to capture true correlations. Although these methods are well-founded in mathematics, have been benchmarked and validated in only limited circumstances and their performance has not yet been characterized for metabolomic data in general.

**Conclusion**

The overall diversity and plasticity of the gut microbiota, in comparison to our human genomes, provides exciting new prospects for personalized medicine—particularly for studies to determine the mechanisms by which microbes affect production of metabolites from drugs and diet. Although there is much work to be done, especially in terms of computational methods, the experimental frameworks of metabolomics and microbial community analysis that have emerged should allow for rapid host characterization followed by subsequent analyses of clinical potential.

# Chapter VI

# Potato Metabolic Profile Variability

This chapter highlights work I lead in collaboration with Prof. Jairam Vanamala at Pennsylvania State University. As I described in Chapter V, knowledge of the metabolome is critical for interpreting changes in the microbiome and the eventual effect on the host. Therefore, we sought to provide a comprehensive picture of the diversity and range of the metabolite profile in a common food. Potatoes represent one of the most abundant and ubiquitous foods to western society; however, researchers do not have a full appreciation for the impact that various processing methods and cultivar strains might have on the overall metabolomics profile. This work assesses how the overall metabolite profile and the relative concentration of common vitamins, minerals, and glycoalkaloids are affected across 6 popular processing methods and 6 common potato cultivars.

## Introduction

Cooking foods, especially starch-rich tubers, represents a uniquely human activity that allows us to extract more energy from our diet due to improved digestibility [235]. As such, the advent of cooking foods and the resulting improvement of energy utilization is hypothesized to have played a major role in human evolution [236]. Potatoes (Solanum tuberosum) represent a ubiquitous and important nutrient-dense component of many diets around the world. However, the

cultivars of potatoes consumed as well as the manner in which potatoes are processed vary greatly from region to region and are likely to influence potato nutritional content [237, 238]. We utilized untargeted metabolomics, targeted metabolomics, and specific assays to assess the changes in important vitamins, antioxidants and toxicants as a function of potato cultivar and processing method. Six different potato types were chosen, two white-, red-, and purple-fleshed cultivars; each was processed 6 different ways and prepared for metabolite analysis (**Table 6.1, Appendix C**).

**Metabolomic Profile Dominated by Processing Method**

Untargeted metabolomics using tandem mass spectrometry was performed on all the 36 treatment groups (six cultivars x six processing methods), yielding 15,129 shared metabolites (see **Appendix C**). We calculated Bray-Curtis distances on the log transformed intensities and applied the dimensionality-reduction technique of Principal Coordinates Analysis (PCoA) to visualize the distances. The clustering in PCoA space utilizing all metabolites showed that processing method resulted in the most delineated separation, while clustering by potato cultivar was observed within each processing method (**Figure 6.1a, b**). PERMANOVA revealed that 80% of the variation in metabolites was the result of processing method, and 9.3% the result of potato cultivar (**Table S6.1**).

**Overview of processing methods and potato cultivars**

| Processing Method | Processing Method Descriptions |
|---|---|
| Baked | Tubers wrapped in aluminum foil were baked at 204°C for one hour in a conventional oven. Tuber internal temperature reached ~ 95 ± 5°C. |
| Chipped | Chip slices were cooked in oil at 185°C for 2 minutes. |
| Fried | Tuber strips were cooked in oil at 190°C for 3.5 minutes. |
| Microwaved | Tubers were cooked for 3.5 min in a 1100 watt microwave oven. Tuber internal temperature reached up to ~ 91 ± 4°C. |
| Raw | Not cooked |
| Steamed | Tubers were placed over boiling water in a steamer for 30 minutes. Tuber internal temperature reached up to ~ 84 ± 3°C. |

**Potato Cultivars**

All Blue | Purple Majesty | All Red | Mountain Rose | Atlantic | Russet Burbank

Purple-fleshed | Red-fleshed | White-fleshed

**Table 6.1. Overview of processing methods and potato cultivars.** Six different potato cultivars and six different processing methods were selected. Each possible combination of potato cultivar by processing method (Atlantic-Raw, Atlantic-Baked, etc.) was performed (36 combinations) in triplicate (n = 108 samples total). Two different potato cultivars from purple-, red-, and white-fleshed potatoes were used.

Similarly, across all metabolites ANOSIM revealed significant differences between processing methods and potato flesh-colors, but differences resulting from individual potato cultivars were not significant after correcting for multiple comparisons (**Tables S6.2-S6.4**). The large effect of processing method on metabolites prompted us to investigate the relative variability of metabolites within each potato cultivar (**Tables S6.5, S6.6**). Procrustes analysis (**Figure 6.1c)**, where the orientation of points in PCoA space is scaled and rotated to minimize the distances between the different processing methods within each cultivar, revealed that all six potato cultivars underwent marked changes in metabolite profile as a result of cooking. However, the orientation of these changes differed with potato cultivar. No one potato cultivar had a processed metabolite profile that looked more similar to their uncooked, raw metabolite profile compared to other cultivars.

**Processing – Cultivar Interaction**

The variability in the overall metabolites caused by processing method and potato cultivar (**Figure 6.1c**) led us to believe that these two variables might interact non-linearly. An additive model of the two variables (processing method and potato cultivar) would result in the abundance of a given metabolite being affected by a change in the other variable. An interaction model would result in the abundance of a metabolite in one variable being dependent on the level in the other variable (see example interaction plots of the additive and interaction model in **Figure S6.2**).

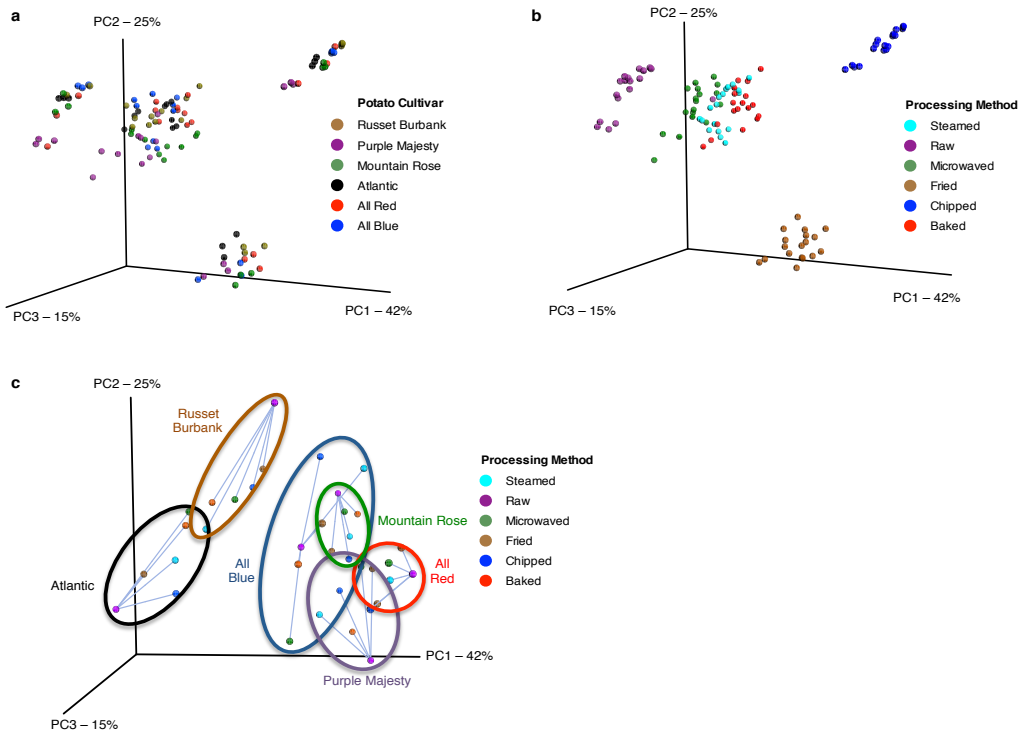


**Figure 6.1**. **Bray-Curtis distances visualized with principal coordinates analysis of 15,129 metabolites reveals the strong effect of processing method compared to potato cultivar on metabolite profile**. Each point on the graph represents the entire metabolic profile of a single replicate. There are 3 replicates for each combination of potato cultivar by processing method. Points that are closer to each other in PCoA space have a more similar metabolic profile. **b** PCoA colored by processing method. The largest variation is explained along PC1 and is driven by the chipped and fried processing of potatoes. PC2 mainly separates out chipped from fried. Raw potatoes cluster towards one end of PC3 compared to microwaved, steamed, and baked potatoes (also see **Table S6.3**). **b** PCoA colored by potato cultivar. No clear trends of clustering by potato cultivar are seen within each processing method (also see **Tables S6.2, S6.4, S6.5**). **c** Procrustes plot of averaged metabolites. Each node represents the 3-replicate average of a processed potato cultivar. The drawn ellipticals enclose all samples from a single potato type, while each colored node located within an elliptical represents a different processing method. The purple centroids of the connected clusters represent the raw potatoes, which are connected to the chipped, fried, microwaved, baked, and steamed processed potatoes of the same potato cultivar with a light blue line. The distance between the raw centroid and each processing method point was calculated for each potato cultivar, and revealed that no cultivar had processed samples that more closely resembled their raw counterparts than any other cultivar (Kruskal Wallis, p-value > 0.05).

We therefore used the two-way ANOVA test to determine if the variables of potato cultivar and processing method demonstrate an interaction effect or if these variables were additive across all 15,129 metabolites. This analysis revealed that ~73% of all the potato product metabolites had relative concentrations that were significantly altered by potato cultivar, processing method, and an interaction effect (**Figure S6.1**).

Because the complete metabolite profile of potatoes differed depending on the interaction between potato cultivar and processing method, we tested for similar patterns in important antioxidants and vitamins common to potatoes. The relative levels of caffeic acid, chlorogenic acid, rutin, genistein, cyanidin, petunidin, caffeoylcholine and trigonelline were significantly affected by potato cultivar, processing method and interaction (**Figure S6.2**). However, folic acid (vitamin B9) was only affected by processing and interaction, and not potato cultivar. The common anthocyanidins, including cyanidin and petunidin, were significantly higher in the color-fleshed cultivars compared to the white-fleshed potatoes (**Figure S6.2**).

**Variability in Healthy and Unhealthy Compounds**

The potato is an important dietary source of vitamin C in the US and many parts of the world [237]. Indeed, potatoes are characterized by government agencies as an affordable source of the daily recommended intake of vitamin C[239, 240]. Therefore, we analyzed the vitamin C levels via high performance liquid

chromatography to provide absolute concentrations (see **Appendix C, Methods**). Like most metabolites, vitamin C levels were significantly altered by potato cultivar, processing method, and by interaction (two-way ANOVA, $p < 0.05$, FDR corrected). The levels of vitamin C were similar in the raw potatoes of each cultivar, yet the levels varied differentially based on both processing method and cultivar (**Figure 6.2**). For example, baked white-fleshed Atlantic potato retained twice the vitamin C levels of baked Purple Majesty potatoes. However, chipping and frying drastically reduced vitamin C across all potato cultivars. Our results indicated that no potato cultivar, and no processing method, systematically produced the highest relative concentrations of all beneficial metabolites. However, we wanted to establish which potato cultivars and processing method generally lead to the highest values for the vitamins and antioxidants listed above. Given that range of each metabolite concentration might be different, the concentrations were transformed into ranks (lower values having lower ranks) to allow direct comparison. We averaged together all ranks for a given potato flesh color across all vitamins and antioxidants. Both the purple-fleshed potatoes (Purple Majesty and All Blue) and the red-fleshed potatoes (All Red and Mountain Rose) had significantly higher average ranks of these beneficial metabolites than the white-fleshed potatoes (Russet Burbank and Atlantic) across all processing methods (comparing average ranks between potatoes of different flesh colors, Student's t-test, unequal variance, $p << 0.01$, $p << 0.01$, respectively), while the purple- and red-fleshed potatoes were comparable (**Figure 6.2b)**.

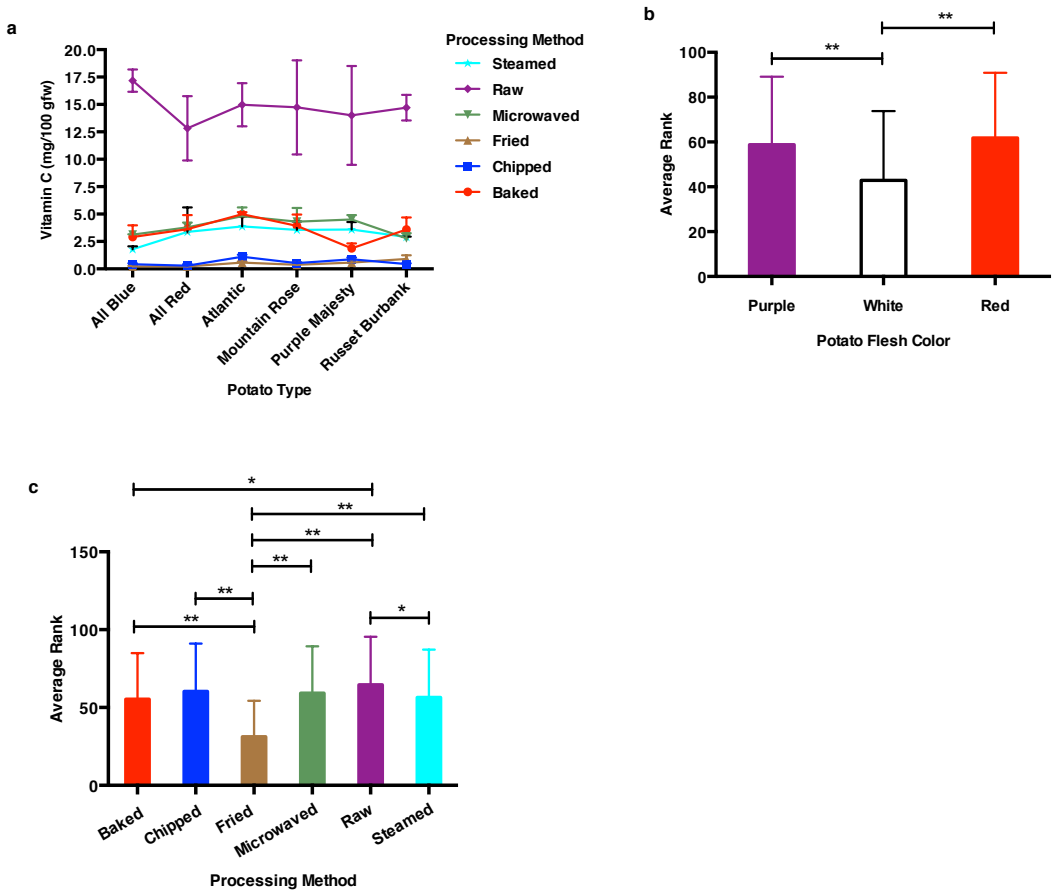


**Figure 6.2. Effect of processing method on vitamin C, and average vitamin and antioxidant rank across flesh-colors and processing methods. a** The plot indicates the average (n = 3 replicates) vitamin C content (mg/100 gram fresh weight; gfw) for each treatment group, with error bars representing ± standard deviation. **b** The average rank (lower values having lower ranks) across nine metabolites (caffeic acid, chlorogenic acid, rutin, genistein, cyanidin, caffeoylcholine, trigonelline, folic acid, and vitamin C) was calculated for each potato flesh-color. Over these metabolites, purple-fleshed potatoes and red-fleshed potatoes had a significantly higher average rank compared to white-fleshed potatoes. **c** Similarly, the average rank across these nine metabolites was calculated for each cooking method across all six potato cultivars. French fried potatoes resulted in the lowest average rank across these metabolites compared against all other processing methods. * indicates p-values < 0.05, ** indicates p-values << 0.05, Student's t-test.

We further assessed which cooking method, irrespective of potato cultivar, yielded the highest overall vitamin and antioxidant rank. Similar to the results obtained from the two-way ANOVA tests, frying resulted in a significantly lower rank across these vitamins and antioxidants compared to all other processing methods (**Figure 6.2c**). Taking into account both potato cultivar and processing method across these select nine vitamins and antioxidants, a fried Russet Burbank potato resulted in the lowest rank while a steamed All Red resulted in the highest (Student's t-test, $p = 0.0002$). The general pattern was for purple- and red-fleshed potatoes to be the higher in vitamins and antioxidants than white-fleshed potatoes within each processing method. However, due to interaction, notable exceptions occurs, such as the Atlantic baked potato having twice the Vitamin C as a Purple Majesty baked potato seen above.

Despite the presence of many beneficial compounds, potatoes also contain some harmful compounds. The glycoalkaloids solanine and chaconine are known toxins, which may produce harmful effects in humans [241]. Similarly, acrylamide has been identified as potential carcinogen in potatoes that results from high heat cooking methods, particularly baking, frying, and chipping [242, 243]. We therefore assessed the range of concentrations and the interaction effect in potato cultivars and processing methods for these compounds (**Figure 6.3**).

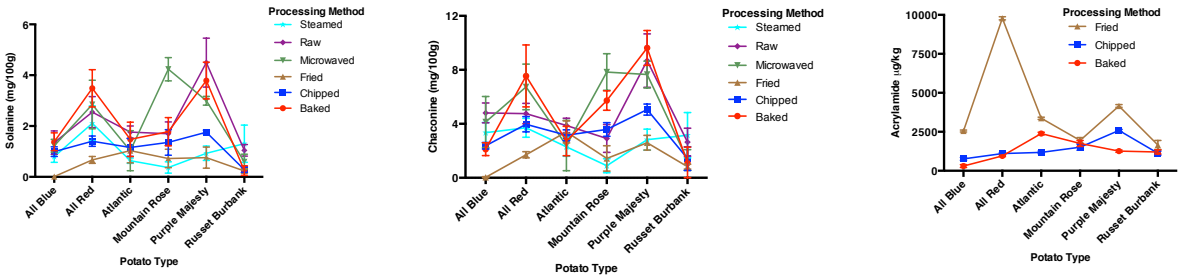


**Figure 6.3. Concentrations of glycoalkaloids (solanine and chaconine) and acrylamide across potato cultivars and processing methods.** The plot indicates the average glycoalkaloids (mg/100 gram fresh weight) solanine (**a**), chaconine (**b**), and acrylamide (**c**) concentrations for n = 3 replicates, with error bars representing ± their standard deviations. Acrylamide was undetectable in raw, steamed, and microwaved potatoes. All three compounds were significantly affected by potato cultivar, processing method and interaction term at p-value << 0.01, two-way ANOVA.

All three harmful compounds were significantly affected by potato cultivar, processing method, and interaction (two-way ANOVA, $p << 0.01$). Solanine and chaconine demonstrated very similar changes across cultivars and processing methods. Acrylamide was not detected in raw, steamed, or microwaved potatoes, which is consistent with previous results, and was highest in French fried potatoes, which aligns with increased acrylamide as a function of high heat and cooking time [242]. The potato cultivars of All Red and Purple Majesty, while generally higher in vitamins and antioxidants, were also higher in glycoalkaloids (however, below FDA approved safe limit of total glycoalkaloids 20 mg/100 gram fresh weight) when processed via baking.

**Discussion**

The findings of this work have broad implications for food science, public health and the average consumer. Firstly, most food science studies assess the change in nutrients across only one variable, either cultivar type or processing method, with the assumption that the two variables abide by the additive model. However, this work reveals that both variables must be evaluated experimentally to gain an understanding of the range of concentrations of important vitamins, antioxidants and toxicants, because of the strong interaction effect. We have previously shown that raw and baked potatoes of the Purple Majesty cultivar maintained their anti-proliferative and pro-apoptotic properties in both early and advanced human colon cancer cells, while chipping this same purple-fleshed potato

resulted in a significant decrease in these benefits [244]. This work tempers those findings by demonstrating that pro-apoptotic properties, while present in some raw and baked potatoes, cannot be assumed to extend to other cultivars due to interaction.

These results further demonstrate that dietary recommendations and food questionnaires should include both the cultivar and processing method employed. For example, proponents of potatoes in school lunches cite that potatoes are an affordable and dense source of vitamin C[237], yet the most frequently served potato cultivars are white-fleshed and are usually fried. Our data indicate that these fried potatoes exhibit the lowest vitamin C and nutrient concentrations and the highest levels of acrylamide. Therefore, school lunches should utilize specific potato cultivars and processing methods if the goal is to provide targeted vitamin C levels. The Institute of Medicine has put forth Dietary Reference Intakes (DRIs) as a guideline for nutrient reference values for more than 40 nutrient substances, and these values are the basis for recommendations in public policies and programs, as well as from individual health care practitioners[1]. This study demonstrates the challenge in accurately providing specific food recommendations in the setting of varied cultivars and cooking methods. Further, the results of this study highlight the potential for misrepresentation of nutrient contents on Nutrition Facts labels on food products to consumers. Though the FDA provides guidance on labeling for food manufacturers, it does not specifically stipulate how nutrient contents should be determined, and the usage of average values from databases is allowed[245].

However, as mass spectrometry techniques continue to decline dramatically in price and increase in throughput, actual measurements of the cooked food product may replace database values that may not be relevant.

Additionally, our data demonstrate that the nutritional benefits of tubers to early humans were likely heavily dependent on both the type of tuber being consumed and the way in which it was processed. This interaction effect suggests that the nutritional consequences of incorporating cooked tubers into their diets may have been more complicated than currently assumed. Although the utilization of some varieties of tubers may have paved the way for large brains and expanded ranges [246, 247], the utilization of others may have reduced vitamin intake and increased toxin intake. Consequently, food selectivity was likely critical despite the apparent dietary flexibility cooking provides.

The net effect of processing and cultivar on the host is not currently known. Cooking can not only alter the micronutrient and antioxidant compound levels, but may also have a differential effect on digestibility of plant foods, chances of escaping digestion in the upper gastrointestinal tract, and being metabolized by gut microbes [188, 248, 249]. Thus, the assessment of the effects of farm-to-fork operations on the net health benefit of foods to the host requires quantitative analytical techniques and the selection of appropriate in vivo models.

# Chapter VII

## Conclusions

The rapid development of next-generation sequencing technologies over the past decade has allowed for the characterization of the bacteria that live in and on the human body at unprecedented levels. Early survey studies that harnessed this sequencing discovered the amazing diversity and uniqueness of every individual's microbiome, and how the microbiome was associated with alterations in a number of disease states. However, researchers also discovered that the microbiome was a critical factor to be assessed for drug targets and medicinal interventions.

While the definition of a "healthy" microbiome in various body habitats has been a much-debated topic, there are several microbial community structures that have been associated with disease states. A well-characterized deleterious microbial community is involved in bacterial vaginosis, whereby a typically health vaginal community that is dominated by a single Lactobacillus member is instead characterized as having a far more heterogeneous microbial membership. This fact is important to consider, as any medical interventions that cause the vaginal community to adopt a more "disease-like" microbial community may likely result in complications.

Therefore, I assessed whether the use of intravaginal rings for localized delivery of acyclovir medication in women with recurrent herpes simplex virus was causing any significant alterations in the patient's vaginal communities. By utilizing 16S rRNA sequencing, I was able to characterize their vaginal microbial

communities even though the majority of members are not culturable. The study design also included a time-series component, whereby repeated measurements were taken in the same individual over time. My time-series analysis demonstrated that the implantation of the intravaginal rings did not significantly alter each patient's vaginal microbial community. No microbes were enriched or depleted as a result of the ring's presence. Additionally, the biofilm that formed on the intravaginal rings was statistically identical to the patient's overall vaginal microbial community, demonstrating the safety of the device. This study serves as an important reminder that medical interventions must not only demonstrate efficacy in their intended target and purpose, but also that the interventions must not alter the microbiome in harmful ways.

While the above situation describes how a defined intervention could alter the microbiome, it is much harder to assess how our current cultural lifestyles might be shifting our microbiome with respect to our ancestors. The analysis of the microbiome from individuals in traditional societies, especially hunter-gatherers, gives us an insight into our microbial past. We can evaluate the membership and functional capacity of our microbiome from a time when humans were hunting and gathering in groups, far before the invention of farming and the agrarian lifestyle. My analysis of the Matses hunter-gatherer peoples demonstrates that while they harbor a low-diversity gut community, their microbes are enriched for genes associated with metabolism and energy extraction from their diets, which likely confirmed an evolutionary survival benefit. Its interesting that they same low-

diversity microbial community structure has been associated with obesity in modern day populations, but the Matses are not overweight.

This dissertation also summarizes many modern-day practices that seem to have altered our microbial communities away from their ancestral, evolutionary paths by reducing our exposure to beneficial microbes, including exposure to animals and livestock, the use of antibiotics, and the use of cesarean sections. Each of these practices have been linked with alterations in the gut microbiome that may be affecting our health, notably by reducing our exposure to microbes that train our immune system to recognize beneficial bacteria, and therefore contributing to conditions such as allergies. This overall series of events has been termed the "hygiene hypothesis", where modern society's obsession with "cleanliness" is actually harming us by reducing our exposure to beneficial microbes. However, researchers are developing new ways of introducing back into our microbiome helpful bacteria through mechanisms such as fecal transplants, probiotics, and prebiotics.

Another interesting facet of the microbiome that this dissertation explores is the interaction between the host microbiome and metabolomics. Humans consume a variety of compounds, including food and xenobiotics such as pharmaceutical drugs, which become substrates for our gut microbiota. Different members of the microbiota can utilize the available substrates in different ways, causing a variety of secondary metabolites to be formed, sometimes with unintended consequences. For instance, pharmaceutical drugs can be inactivated by the presence of certain

microbial members, who metabolize necessary functional groups. Additionally, microbes can produce harmful secondary metabolites from common drugs such as acetaminophen. While the need to integrate microbiome and metabolomics datasets is obvious, the computational challenges are still significant. However, computational and algorithmic tools are constantly improving, and the future of multi-omic data integration is promising.

After summarizing the need to consider both the microbiome and metabolome when assessing dietary interventions, I lead analysis on a collaboration that evaluated changes in the overall metabolite profile as well as changes in specific vitamins, minerals, and glycoalkaloids in potatoes across different process methods and cultivars. Remarkably, we demonstrated that processing method and cultivar interact to produce unpredictable concentrations in the relative abundances of the majority of metabolites. This means that knowledge of the relative concentration of a metabolite cannot be extrapolated from the knowledge of the metabolites concentration in that processing method and in that cultivar, because the interaction produces a non-linear combination of variables. We also demonstrated that in general purple-fleshed potatoes harbor higher concentrations of healthy vitamins and antioxidants than their white-fleshed counterparts. We also showed that processing method was the strongest determinant of the overall metabolic profile, regardless of cultivar type. Taken together, these findings necessitate that food researchers cannot estimate a food's concentration of important metabolites without directly testing the combination of processing

method and cultivar. It also suggests that these variables can be manipulated in the future to produce foods that maximize the relative concentrations of desirable compounds while minimizing the concentrations of harmful ones. Future work will include feeding these potato-containing diets to animal models to assess if the vastly different metabolite profiles can globally affect markers of host health.

Taken together, this dissertation demonstrates the power of both manipulating the human microbiome and the importance in assessing how the human microbiome is changing as a result of dietary inputs and medical interventions.

# BIBLIOGRAPHY


1. Institute of Medicine (U.S.). Panel on Macronutrients., Institute of Medicine (U.S.). Standing Committee on the Scientific Evaluation of Dietary Reference Intakes.: *Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein, and amino acids.* Washington, D.C.: National Academies Press; 2005.

2. Ursell LK, Clemente JC, Rideout JR, Gevers D, Caporaso JG, Knight R: **The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites.** *J Allergy Clin Immunol* 2012, **129:**1204-1208.

3. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI: **The human microbiome project.** *Nature* 2007, **449:**804-810.

4. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, et al: **The NIH Human Microbiome Project.** *Genome Research* 2009, **19:**2317-2323.

5. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* 2010, **464:**59-65.

6. Lederberg J, McCray A: **Ome sweet 'omics: -- A genealogical treasury of words.** *The Scientist* 2001, **15:**8.

7. Dobell C: **The discovery of intestinal protozoa in man.** *Proceedings of the Royal Society of Medicine* 1920, **13:**1-15.

8. van Leeuwenhoek A: **An abstract of a Letter from Antonie van Leeuwenhoek,Sep. 12, 1683. About Animals in the scrurf of the Teeth.** *Philosophical Transactions of the Royal Society of London* 1684, **14:**568-574.

9. Pace NR: **A molecular view of microbial diversity and the biosphere.** *Science* 1997, **276:**734-740.

10. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI: **Worlds within worlds: evolution of the vertebrate gut microbiota.** *Nature Reviews Microbiology* 2008, **6:**776-788.

11. Lozupone CA, Knight R: **Global patterns in bacterial diversity.** *Proc Natl Acad Sci U S A* 2007, **104:**11436-11440.

12. Marchesi JR: **Prokaryotic and Eukaryotic Diversity of the Human Gut.** In *Advances in Applied Microbiology, Vol 72. Volume* 722010: 43-62

13. Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, Mahaffy JM, Mueller J, Nulton J, Rayhawk S, et al: **Viral diversity and dynamics in an infant gut.** *Research In Microbiology* 2008, **159:**367-373.

14. Consortium IHGS: **Finishing the euchromatic sequence of the human genome.** *Nature* 2004, **431:**931-945.

15. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, et al: **The complete genome of an individual by massively parallel DNA sequencing.** *Nature* 2008, **452:**872-876.

16. Fierer N, Hamady M, Lauber CL, Knight R: **The influence of sex, handedness, and washing on the diversity of hand surface bacteria.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105:**17994-17999.

17. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, et al: **A core gut microbiome in obese and lean twins.** *Nature* 2009, **457:**480-484.

18. Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D, Knight R: **Experimental and analytical tools for studying the human microbiome.** *Nature reviews Genetics* 2012, **13:**47-58.

19. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al: **Introducing mothur: open-source, platform-independent, community-supported**

**software for describing and comparing microbial communities.** *Applied and environmental microbiology* 2009, **75:**7537-7541.

20. Hartman AL, Riddle S, McPhillips T, Ludascher B, Eisen JA: **Introducing W.A.T.E.R.S.: a workflow for the alignment, taxonomy, and ecology of ribosomal sequences.** *BMC bioinformatics* 2010, **11:**317.

21. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM: **The Ribosomal Database Project: improved alignments and new tools for rRNA analysis.** *Nucleic acids research* 2009, **37:**D141-145.

22. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al: **QIIME allows analysis of high-throughput community sequencing data.** *Nature Methods* 2010, **7:**335-336.

23. Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, Knight R: **Microbial community resemblance methods differ in their ability to detect biologically relevant patterns.** *Nat Methods* 2010, **7:**813-819.

24. Lozupone C, Knight R: **UniFrac: a new phylogenetic method for comparing microbial communities.** *Applied and environmental microbiology* 2005, **71:**8228-8235.

25. Kuczynski J, Costello EK, Nemergut DR, Zaneveld J, Lauber CL, Knights D, Koren O, Fierer N, Kelley ST, Ley RE, et al: **Direct sequencing of the human microbiome readily reveals community differences.** *Genome Biology* 2010, **11**.

26. Muegge BD, Kuczynski J, Knights D, Clemente JC, Gonzalez A, Fontana L, Henrissat B, Knight R, Gordon JI: **Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans.** *Science* 2011, **332:**970-974.

27. Knights D, Costello EK, Knight R: **Supervised classification of human microbiota.** *FEMS microbiology reviews* 2011, **35:**343-359.

28. White JR, Nagarajan N, Pop M: **Statistical methods for detecting differentially abundant features in clinical metagenomic samples.** *PLoS computational biology* 2009, **5:**e1000352.

29. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, et al: **Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications.** *Nature biotechnology* 2011, **29:**415-420.

30. Yilmaz P, Gilbert JA, Knight R, Amaral-Zettler L, Karsch-Mizrachi I, Cochrane G, Nakamura Y, Sansone SA, Glockner FO, Field D: **The genomic standards consortium: bringing standards to life for microbial ecology.** *The ISME journal* 2011.

31. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, et al: **The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes.** *BMC Bioinformatics* 2008, **9:**386.

32. Mackie RI, Sghir A, Gaskins HR: **Developmental microbial ecology of the neonatal gastrointestinal tract.** *Am J Clin Nutr* 1999, **69:**1035S-1045S.

33. Dominguez-Bello MG, Blaser MJ, Ley RE, Knight R: **Development of the human gastrointestinal microbiota and insights from high-throughput sequencing.** *Gastroenterology* 2011, **140:**1713-1719.

34. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R: **Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107:**11971-11975.

35. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO: **Development of the human infant intestinal microbiota.** *PLoS Biol* 2007, **5:**e177.

36. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE: **Succession of microbial consortia in the**

**developing infant gut microbiome.** *Proceedings of the National Academy of Science, USA* 2010.

37. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R: **Forensic identification using skin bacterial communities.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107:**6477-6481.

38. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, et al: **Moving pictures of the human microbiome.** *Genome Biology* 2011, **12:**R50.

39. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R: **Bacterial Community Variation in Human Body Habitats Across Space and Time.** *Science* 2009, **326:**1694-1697.

40. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P: **Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates.** *Environmental Microbiology* 2010, **12:**118-123.

41. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT: **Accurate determination of microbial diversity from 454 pyrosequencing data.** *Nature Methods* 2009, **6:**639-641.

42. Sun Y, Cai Y, Huse SM, Knight R, Farmerie WG, Wang X, Mai V: **A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis.** *Briefings in bioinformatics* 2011.

43. Schloss PD, Westcott SL: **Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis.** *Applied and Environmental Microbiology* 2011, **77:**3219-3226.

44. Huse SM, Welch DM, Morrison HG, Sogin ML: **Ironing out the wrinkles in the rare biosphere through improved OTU clustering.** *Environmental Microbiology* 2010, **12:**1889-1898.

45. Sun Y, Cai Y, Liu L, Yu F, Farrell ML, McKendree W, Farmerie W: **ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences.** *Nucleic acids research* 2009, **37:**e76.

46. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, et al: **Vaginal microbiome of reproductive-age women.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108 Suppl 1:**4680-4687.

47. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, et al: **Enterotypes of the human gut microbiome.** *Nature* 2011, **473:**174-180.

48. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, et al: **Functional metagenomic profiling of nine biomes.** *Nature* 2008, **452:**629-632.

49. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, et al: **Comparative metagenomics of microbial communities.** *Science* 2005, **308:**554-557.

50. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI: **Evolution of mammals and their gut microbes.** *Science* 2008, **320:**1647-1651.

51. Martin FP, Sprenger N, Montoliu I, Rezzi S, Kochhar S, Nicholson JK: **Dietary modulation of gut functional ecology studied by fecal metabonomics.** *Journal of proteome research* 2010, **9:**5284-5295.

52. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI: **The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice.** *Science translational medicine* 2009, **1:**6ra14.

53. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: **An obesity-associated gut microbiome with increased capacity for energy harvest.** *Nature* 2006, **444:**1027-1031.

54. Gilmore MS, Ferretti JJ: **Microbiology: The thin line between gut commensal and pathogen.** *Science* 2003, **299:**1999-+.

55. Backhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI: **The gut microbiota as an environmental factor that regulates fat storage.** *Proc Natl Acad Sci U S A* 2004, **101:**15718-15723.

56. Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI: **Obesity alters gut microbial ecology.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102:**11070-11075.

57. Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, Sitaraman SV, Knight R, Ley RE, Gewirtz AT: **Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5.** *Science* 2010, **328:**228-231.

58. Ley RE, Turnbaugh PJ, Klein S, Gordon JI: **Microbial ecology: human gut microbes associated with obesity.** *Nature* 2006, **444:**1022-1023.

59. Hehemann J, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G: **Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota.** *Nature* 2010, **464:**908-912.

60. Dethlefsen L, Huse S, Sogin ML, Relman DA: **The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing.** *Plos Biology* 2008, **6:**2383-2400.

61. Dethlefsen L, Relman DA: **Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108 Suppl 1:**4554-4561.

62. Dethlefsen L, Huse S, Sogin ML, Relman DA: **The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing.** *Plos Biology* 2008, **6:**e280.

63. Jernberg C, Lofmark S, Edlund C, Jansson JK: **Long-term ecological impacts of antibiotic administration on the human intestinal**

**microbiota.** *The ISME journal* 2007, **1:**56-66.

64. Jernberg C, Lofmark S, Edlund C, Jansson JK: **Long-term ecological impacts of antibiotic administration on the human intestinal microbiota.** *Isme Journal* 2007, **1:**56-66.

65. Jernberg C, Sullivan A, Edlund C, Jansson JK: **Monitoring of antibiotic-induced alterations in the human intestinal microflora and detection of probiotic strains by use of terminal restriction fragment length polymorphism.** *Applied and Environmental Microbiology* 2005, **71:**501-506.

66. Jakobsson HE, Jernberg C, Andersson AF, Sjolund-Karlsson M, Jansson JK, Engstrand L: **Short-Term Antibiotic Treatment Has Differing Long-Term Impacts on the Human Throat and Gut Microbiome.** *PLoS ONE* 2010, **5**.

67. Manichanh C, Reeder J, Gibert P, Varela E, Llopis M, Antolin M, Guigo R, Knight R, Guarner F: **Reshaping the gut microbiome with bacterial transplantation and antibiotic intake.** *Genome Research* 2010, **20:**1411-1419.

68. Mellow MH, Kanatzar A: **Colonoscopic fecal bacteriotherapy in the treatment of recurrent clostridium difficile infection--results and follow-up.** *The Journal of the Oklahoma State Medical Association* 2011, **104:**89-91.

69. Grehan MJ, Borody TJ, Leis SM, Campbell J, Mitchell H, Wettstein A: **Durable alteration of the colonic microbiota by the administration of donor fecal flora.** *Journal of Clinical Gastroenterology* 2010, **44:**551-561.

70. Garborg K, Waagsbo B, Stallemo A, Matre J, Sundoy A: **Results of faecal donor instillation therapy for recurrent Clostridium difficile-associated diarrhoea.** *Scandinavian journal of infectious diseases* 2010, **42:**857-861.

71. Silverman MS, Davis I, Pillai DR: **Success of self-administered home fecal transplantation for chronic Clostridium difficile infection.** *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2010, **8:**471-473.

72. MacConnachie AA, Fox R, Kennedy DR, Seaton RA: **Faecal transplant for recurrent Clostridium difficile-associated diarrhoea: a UK case series.** *QJM : monthly journal of the Association of Physicians* 2009, **102:**781-784.

73. You DM, Franzos MA, Holman RP: **Successful treatment of fulminant Clostridium difficile infection with fecal bacteriotherapy.** *Annals of internal medicine* 2008, **148:**632-633.

74. Tvede M, Rask-Madsen J: **Bacteriotherapy for chronic relapsing Clostridium difficile diarrhoea in six patients.** *Lancet* 1989, **1:**1156-1160.

75. Khoruts A, Dicksved J, Jansson JK, Sadowsky MJ: **Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea.** *Journal of clinical gastroenterology* 2010, **44:**354-360.

76. Faith JJ, Rey FE, O'Donnell D, Karlsson M, McNulty NP, Kallstrom G, Goodman AL, Gordon JI: **Creating and characterizing communities of human gut microbes in gnotobiotic mice.** *The ISME journal* 2010, **4:**1094-1098.

77. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI: **Identifying genetic determinants needed to establish a human gut symbiont in its habitat.** *Cell Host & Microbe* 2009, **6:**279-289.

78. Ursell LK, Gunawardana M, Chang S, Mullen M, Moss JA, Herold BC, Keller MJ, McDonald D, Gonzalez A, Knight R, Baum MM: **Comparison of the vaginal microbial communities in women with recurrent genital HSV receiving acyclovir intravaginal rings.** *Antiviral Res* 2014, **102:**87-94.

79. Looker KJ, Gamett GP, Schmid GP: **An Estimate of the Global Prevalence and Incidence of Herpes Simplex Virus Type 2 Infection.** *Bull World Health Organ* 2008, **86:**805-812.

80. Corey L, Wald A, Patel R, Sacks SL, Tyring SK, Warren T, Douglas JM, Paavonen J, Morrow RA, Beutner KR, et al: **Once-daily Valacyclovir to Reduce the Risk of Transmission of Genital Herpes.** *N Engl J Med*

2004, **350:**11-20.

81. Freeman EE, Weiss HA, Glynn JR, Cross PL, Whitworth JA, Hayes RJ: **Herpes Simplex Virus 2 Infection Increases HIV Acquisition in Men and Women: Systematic Review and Meta-analysis of Longitudinal Studies.** *AIDS* 2006, **20:**73-83.

82. Freeman EE, Orroth KK, White RG, Glynn JR, Bakker R, Boily MC, Habbema D, Buve A, Hayes RJ: **Proportion of New HIV Infections Attributable to Herpes Simplex 2 Increases over Time: Simulations of the Changing Role of Sexually Transmitted Infections in sub-Saharan African HIV Epidemics.** *Sex Transm Infect* 2007, **83:**I17-I24.

83. Patel R, Bodsworth NJ, Woolley P, Peters B, Vejlsgaard G, Saari S, Gibb A, Robinson J: **Valaciclovir for the Suppression of Recurrent Genital HSV Infection: A Placebo Controlled Study of Once Daily Therapy.** *Genitourin Med* 1997, **73:**105-109.

84. Corey L, Benedetti JK, Critchlow CW, Remington MR, Winter CA, Fahnlander AL, Smith K, Salter DL, Keeney RE, Davis LG, et al: **Double-blind Controlled Trial of Topical Acyclovir in Genital Herpes Simplex Virus Infections.** *Am J Med* 1982, **73:**326-334.

85. Corey L, Nahmias AJ, Guinan ME, Benedetti JK, Critchlow CW, Holmes KK: **A Trial of Topical Acyclovir in Genital Herpes Simplex Virus Infections.** *N Engl J Med* 1982, **306:**1313-1319.

86. Baum MM, Butkyavichene I, Gilman J, Kennedy S, Kopin E, Malone AM, Nguyen C, Smith TJ, Friend DR, Clark MR, Moss JA: **An Intravaginal Ring for the Simultaneous Delivery of Multiple Drugs.** *J Pharm Sci* 2012, **101:**2833-2843.

87. Moss JA, Baum MM, Malone AM, Kennedy S, Kopin E, Nguyen C, Gilman J, Butkyavichene I, Willis R, Vincent KL, et al: **Tenofovir and Tenofovir Disoproxil Pharmacokinetics from Intravaginal Rings.** *AIDS* 2012**:**Epublished ahead of print Jan 4. 2012 (PMID: 22210639).

88. Moss JA, Malone AM, Smith TJ, Kennedy S, Kopin E, Nguyen C, Gilman J, Butkyavichene I, Vincent KL, Motamedi M, et al: **Simultaneous Delivery**

**of Tenofovir and Acyclovir via an Intravaginal Ring.** *Antimicrob Agents Chemother* 2012, **56:**875-882.

89. Moss JA, Malone AM, Smith TJ, Butkyavichene I, Cortez C, Gilman J, Kennedy S, Kopin E, Nguyen C, Sinha P, et al: **Safety and Pharmacokinetics of Intravaginal Rings Delivering Tenofovir in Pig-tailed Macaques.** *Antimicrob Agents Chemother* 2012, **56:**5952-5960.

90. Keller MJ, Malone AM, Carpenter CA, Lo Y, Huang M, Corey L, Willis R, Nguyen C, Kennedy S, Gunawardana M, et al: **Safety and Pharmacokinetics of Acyclovir in Women Following Release From a Silicone Elastomer Vaginal Ring.** *J Antimicrob Chemother* 2012, **67:**2005-2012.

91. Donders GG, Bosmans E, Dekeersmaecker A, Vereecken A, Van Bulck B, Spitz B: **Pathogenesis of abnormal vaginal bacterial flora.** *American journal of obstetrics and gynecology* 2000, **182:**872-878.

92. Gupta K, Stapleton AE, Hooton TM, Roberts PL, Fennell CL, Stamm WE: **Inverse association of H2O2-producing lactobacilli and vaginal Escherichia coli colonization in women with recurrent urinary tract infections.** *The Journal of infectious diseases* 1998, **178:**446-450.

93. Watts DH, Fazzari M, Minkoff H, Hillier SL, Sha B, Glesby M, Levine AM, Burk R, Palefsky JM, Moxley M, et al: **Effects of bacterial vaginosis and other genital infections on the natural history of human papillomavirus infection in HIV-1-infected and high-risk HIV-1-uninfected women.** *The Journal of infectious diseases* 2005, **191:**1129-1139.

94. Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, Joyce P, Foster JA, Forney LJ: **Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women.** *The ISME journal* 2007, **1:**121-133.

95. Ling Z, Kong J, Liu F, Zhu H, Chen X, Wang Y, Li L, Nelson KE, Xia Y, Xiang C: **Molecular analysis of the diversity of vaginal microbiota associated with bacterial vaginosis.** *BMC genomics* 2010, **11:**488.

96. Gunawardana M, Moss JA, Smith TJ, Kennedy S, Kopin E, Nguyen C,

Malone AM, Rabe L, Schaudinn C, Webster P, et al: **Microbial Biofilms on the Surface of Intravaginal Rings Worn in Non-human Primates.** *J Med Microbiol* 2011, **60:**828-837.

97. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, et al: **Ultra-high-throughput Microbial Community Analysis on the Illumina HiSeq and MiSeq Platforms.** *ISME Journal* 2012, **6:**1621-1624.

98. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al: **QIIME Allows Analysis of High-throughput Community Sequencing Data.** *Nat Methods* 2010, **7:**335-336.

99. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG: **Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing.** *Nat Methods* 2013, **10:**57-59.

100. Edgar RC: **Search and Clustering Orders of Magnitude Faster than BLAST.** *Bioinformatics* 2010, **26:**2460-2461.

101. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P: **An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea.** *ISME J* 2012, **6:**610-618.

102. Lladser ME, Gouet R, Reeder J: **Extrapolation of Urn Models via Poissonization: Accurate Measurements of the Microbial Unknown.** *PLoS One* 2011, **6:**e21105.

103. Lozupone CA, Knight R: **Species Divergence and the Measurement of Microbial Diversity.** *FEMS Microbiol Rev* 2008, **32:**557-578.

104. Reichelderfer P: **Toxicology and Safety Testing.** In *The Science of Microbicides: Accelerating Development. Volume* 2. New York: The Rockefeller Foundation; 2002: 37-39

105. Veazey RS: **Microbicide Safety/Efficacy Studies in Animals - Macaques**

**and Small Animal Models.** *Curr Opin HIV AIDS* 2008, **3:**567–573.

106. Van Rompay KKA: **Evaluation of Antiretrovirals in Animal Models of HIV Infection.** *Antiviral Res* 2010, **85:**159-175.

107. Keller MJ, Herold BC: **Understanding Basic Mechanisms and Optimizing Assays to Evaluate the Efficacy of Vaginal Microbicides.** *Sex Transm Dis* 2009, **36:**S92-S95.

108. Ravel J, Gajer P, Fu L, Mauck CK, Koenig SSK, Sakamoto J, Motsinger-Reif AA, Doncel GF, Zeichner SL: **Twice-daily Application of HIV Microbicides Alters the Vaginal Microbiota.** *mBio* 2012, **3**.

109. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, et al: **Vaginal Microbiome of Reproductive-age Women.** *Proc Natl Acad Sci U S A* 2011, **108:**4680-4687.

110. Ursell LK, Clemente JC, Rideout JR, Gevers D, Caporaso JG, Knight R: **The Interpersonal and Intrapersonal Diversity of Human-associated Microbiota in Key Body Sites.** *J Allergy Clin Immunol* 2012, **129:**1204-1208.

111. Kaul R, Nagelkerke NJ, Kimani J, Ngugi E, Bwayo JJ, MacDonald KS, Rebbaprgada A, Fonck K, Temmerman M, Ronald AR, Moses S: **Prevalent Herpes Simplex Virus Type 2 Infection is Associated with Altered Vaginal Flora and an Increased Susceptibility to Multiple Sexually Transmitted Infections.** *J Infect Dis* 2007, **196:**1692-1697.

112. Cherpes TL, Hillier SL, Meyn LA, Busch JL, Krohn MA: **A Delicate Balance: Risk Factors for Acquisition of Bacterial Vaginosis Include Sexual Activity, Absence of Hydrogen Peroxide-producing Lactobacilli, Black Race, and Positive Herpes Simplex Virus Type 2 Serology.** *Sex Transm Dis* 2008, **35:**78-83.

113. Stoner KA, Reighard SD, Miguel RDV, Landsittel D, Cosentino LA, Kant JA, Cherpes TL: **Recalcitrance of Bacterial Vaginosis among Herpes Simplex Virus Type 2 Seropositive Women.** *J Obstet Gynaecol Res* 2012, **38:**77-83.

114. Brotman RM: **Vaginal Microbiome and Sexually Transmitted Infections: an Epidemiologic Perspective.** *J Clin Invest* 2011, **121:**4610-4617.

115. Gajer P, Brotman RM, Bai GY, Sakamoto J, Schuette UME, Zhong X, Koenig SSK, Fu L, Ma ZSS, Zhou X, et al: **Temporal Dynamics of the Human Vaginal Microbiota.** *Sci Transl Med* 2012, **4**.

116. Henzl MR, Mishell DR, Velazque.Jg, Leitch WE: **Basic Studies for Prolonged Progestogen Administration by Vaginal Devices.** *Am J Obstet Gynecol* 1973, **117:**101-106.

117. Aagaard K, Petrosino P, Keitel W, Watson M, Katancik J, Garcia N, Patel S, Cutting M, Madden T, Hamilton H, et al: **The Human Microbiome Project Strategy for Comprehensive Sampling of the Human Microbiome and Why It Matters.** *FASEB J* 2013, **23:**1012-1022.

118. Jackanicz TM: **Levonorgestrel and Estradiol Release from an Improved Contraceptive Vaginal Ring.** *Contraception* 1981, **24:**323-339.

119. Sivin I, Mishell DR, Victor A, Diaz S, Alvarezsanchez F, Nielsen NC, Akinla O, Pyorala T, Coutinho E, Faundes A, et al: **A Multi-center Study of Levonorgestrel-estradiol Contraceptive Vaginal Rings .1. Use Effectiveness - An International Comparative Trial.** *Contraception* 1981, **24:**341-358.

120. Sivin I, Mishell DR, Victor A, Diaz S, Alvarezsanchez F, Nielsen NC, Akinla O, Pyorala T, Coutinho E, Faundes A, et al: **A Multi-Center Study of Levonorgestrel-Estradiol Contraceptive Vaginal Rings .2. Subjective and Objective Measures of Effects - An International Comparative Trial.** *Contraception* 1981, **24:**359-376.

121. Schwan A, Ahren T, Victor A: **Effects of Contraceptive Vaginal Ring Treatment on Vaginal Bacteriology and Cytology.** *Contraception* 1983, **28:**341-347.

122. Davies GC, Feng LX, Newton JR, Dieben TOM, Coelinghbennink HJT: **The Effects of A Combined Contraceptive Vaginal Ring Releasing Ethinylestradiol and 3-Ketodesogestrel on Vaginal Flora.**

*Contraception* 1992, **45:**511-518.

123. Tringe SG, Hugenholtz P: **A Renaissance for the Pioneering 16S rRNA Gene.** *Curr Opin Microbiol* 2008, **11:**442-446.

124. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R: **Global Patterns of 16S rRNA Diversity at a Depth of Millions of Sequences per Sample.** *Proc Natl Acad Sci U S A* 2011, **108:**4516-4522.

125. Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E: **The Use of Coded PCR Primers Enables High-throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing.** *PLoS One* 2007, **2**.

126. Lozupone CA, Hamady M, Kelley ST, Knight R: **Quantitative and Qualitative Beta Diversity Measures Lead to Different Insights into Factors that Structure Microbial Communities.** *Appl Environ Microbiol* 2007, **73:**1576-1585.

127. Kuczynski J, Costello EK, Nemergut DR, Zaneveld J, Lauber CL, Knights D, Koren O, Fierer N, Kelley ST, Ley RE, et al: **Direct Sequencing of the Human Microbiome Readily Reveals Community Differences.** *Genome Biol* 2010, **11**.

128. Ursell LK, Van Treuren W, Metcalf JL, Pirrung M, Gewirtz A, Knight R: **Replenishing our defensive microbes.** *Bioessays* 2013, **35:**810-817.

129. Blaser MJ, Falkow S: **What are the consequences of the disappearing human microbiota?** *Nature reviews Microbiology* 2009, **7:**887-894.

130. Human Microbiome Project C: **Structure, function and diversity of the healthy human microbiome.** *Nature* 2012, **486:**207-214.

131. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* 2010, **464:**59-65.

132. Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ: **Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont.** *Science* 2010, **329:**212-215.

133. Willing BP, Vacharaksa A, Croxen M, Thanachayanont T, Finlay BB: **Altering host resistance to infections through microbial transplantation.** *PLoS ONE* 2011, **6:**e26988.

134. Lawley TD, Clare S, Walker AW, Stares MD, Connor TR, Raisen C, Goulding D, Rad R, Schreiber F, Brandt C, et al: **Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice.** *Plos Pathogens* 2012, **8:**e1002995.

135. Dillon RJ, Vennard CT, Buckling A, Charnley AK: **Diversity of locust gut bacteria protects against pathogen invasion.** *Ecology Letters* 2005, **8:**1291-1298.

136. Havstad S, Wegienka G, Zoratti EM, Lynch SV, Boushey HA, Nicholas C, Ownby DR, Johnson CC: **Effect of prenatal indoor pet exposure on the trajectory of total IgE levels in early childhood.** *J Allergy Clin Immunol* 2011, **128:**880-885 e884.

137. Scanlan PD, Marchesi JR: **Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces.** *ISME J* 2008, **2:**1183-1193.

138. Clemente JC, Ursell LK, Parfrey LW, Knight R: **The impact of the gut microbiota on human health: an integrative view.** *Cell* 2012, **148:**1258-1270.

139. Walter J, Ley RE: **The Human Gut Microbiome: Ecology and Recent Evolutionary Changes.** *Annual review of microbiology* 2010.

140. Lee YK, Mazmanian SK: **Has the microbiota played a critical role in the evolution of the adaptive immune system?** *Science* 2010, **330:**1768-1773.

141. Littman DR, Pamer EG: **Role of the commensal microbiota in normal and pathogenic host immune responses.** *Cell Host & Microbe* 2011, **10:**311-323.

142. Josefowicz SZ, Niec RE, Kim HY, Treuting P, Chinen T, Zheng Y, Umetsu DT, Rudensky AY: **Extrathymically generated regulatory T cells control mucosal TH2 inflammation.** *Nature* 2012, **482:**395-399.

143. Izcue A, Coombes JL, Powrie F: **Regulatory T cells suppress systemic and mucosal immune activation to control intestinal inflammation.** *Immunol Rev* 2006, **212:**256-271.

144. Geuking MB, Cahenzli J, Lawson MA, Ng DC, Slack E, Hapfelmeier S, McCoy KD, Macpherson AJ: **Intestinal bacterial colonization induces mutualistic regulatory T cell responses.** *Immunity* 2011, **34:**794-806.

145. Round JL, Mazmanian SK: **Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota.** *Proc Natl Acad Sci U S A* 2010, **107:**12204-12209.

146. Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y, Cheng G, Yamasaki S, Saito T, Ohba Y, et al: **Induction of colonic regulatory T cells by indigenous Clostridium species.** *Science* 2011, **331:**337-341.

147. Rolig AS, Cech C, Ahler E, Carter JE, Ottemann KM: **The Degree of Helicobacter pylori-Triggered Inflammation Is Manipulated by Preinfection Host Microbiota.** *Infect Immun* 2013, **81:**1382-1389.

148. Lamouse-Smith ES, Tzeng A, Starnbach MN: **The intestinal flora is required to support antibody responses to systemic immunization in infant and germ free mice.** *PLoS One* 2011, **6:**e27662.

149. Ichinohe T, Pang IK, Kumamoto Y, Peaper DR, Ho JH, Murray TS, Iwasaki A: **Microbiota regulates immune defense against respiratory tract influenza A virus infection.** *Proc Natl Acad Sci U S A* 2011, **108:**5354-5359.

150. Johnson PT, Preston DL, Hoverman JT, Richgels KL: **Biodiversity**

**decreases disease through predictable changes in host community competence.** *Nature* 2013, **494:**230-233.

151. Comito D, Romano C: **Dysbiosis in the pathogenesis of pediatric inflammatory bowel diseases.** *International journal of inflammation* 2012, **2012:**687143.

152. Ivanov, II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee CA, Lynch SV, et al: **Induction of intestinal Th17 cells by segmented filamentous bacteria.** *Cell* 2009, **139:**485-498.

153. Ivanov, II, Littman DR: **Modulation of immune homeostasis by commensal bacteria.** *Current opinion in microbiology* 2011, **14:**106-114.

154. Naik S, Bouladoux N, Wilhelm C, Molloy MJ, Salcedo R, Kastenmuller W, Deming C, Quinones M, Koo L, Conlan S, et al: **Compartmentalized control of skin immunity by resident commensals.** *Science* 2012, **337:**1115-1119.

155. Tito RY, Knights D, Metcalf J, Obregon-Tito AJ, Cleeland L, Najar F, Roe B, Reinhard K, Sobolik K, Belknap S, et al: **Insights from characterizing extinct human gut microbiomes.** *PLoS ONE* 2012, **7:**e51146.

156. Tito RY, Macmil S, Wiley G, Najar F, Cleeland L, Qu C, Wang P, Romagne F, Leonard S, Ruiz AJ, et al: **Phylotyping and functional analysis of two ancient human microbiomes.** *PLoS ONE* 2008, **3:**e3703.

157. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, et al: **Human gut microbiome viewed across age and geography.** *Nature* 2012, **486:**222-227.

158. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE: **Succession of microbial consortia in the developing infant gut microbiome.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108 Suppl 1:**4578-4585.

159. Ubeda C, Taur Y, Jenq RR, Equinda MJ, Son T, Samstein M, Viale A, Socci

ND, van den Brink MR, Kamboj M, Pamer EG: **Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans.** *The Journal of clinical investigation* 2010, **120:**4332-4341.

160. Sommer MO, Dantas G, Church GM: **Functional characterization of the antibiotic resistance reservoir in the human microflora.** *Science* 2009, **325:**1128-1131.

161. Looft T, Johnson TA, Allen HK, Bayles DO, Alt DP, Stedtfeld RD, Sul WJ, Stedtfeld TM, Chai B, Cole JR, et al: **In-feed antibiotic effects on the swine intestinal microbiome.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109:**1691-1696.

162. Cho I, Yamanishi S, Cox L, Methe BA, Zavadil J, Li K, Gao Z, Mahana D, Raju K, Teitler I, et al: **Antibiotics in early life alter the murine colonic microbiome and adiposity.** *Nature* 2012, **488:**621-626.

163. Flint HJ: **Microbiology: Antibiotics and adiposity.** *Nature* 2012, **488:**601-602.

164. Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI: **Obesity alters gut microbial ecology.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102:**11070-11075.

165. Trasande L, Blustein J, Liu M, Corwin E, Cox LM, Blaser MJ: **Infant antibiotic exposures and early-life body mass.** *International journal of obesity* 2013, **37:**16-23.

166. Adlerberth I, Wold AE: **Establishment of the gut microbiota in Western infants.** *Acta Paediatr* 2009, **98:**229-238.

167. Huurre A, Kalliomaki M, Rautava S, Rinne M, Salminen S, Isolauri E: **Mode of delivery - effects on gut microbiota and humoral immunity.** *Neonatology* 2008, **93:**236-240.

168. Sjogren YM, Jenmalm MC, Bottcher MF, Bjorksten B, Sverremark-Ekstrom E: **Altered early infant gut microbiota in children developing allergy**

**up to 5 years of age.** *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2009, **39:**518-526.

169. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielny S, et al: **Cohabiting family members share microbiota with one another and with their dogs.** *Elife* 2013, **2:**e00458.

170. Mountzouris KC, McCartney AL, Gibson GR: **Intestinal microflora of human infants and current trends for its nutritional modulation.** *The British journal of nutrition* 2002, **87:**405-420.

171. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R: **Bacterial community variation in human body habitats across space and time.** *Science* 2009, **326:**1694-1697.

172. Turnbaugh PJ, Backhed F, Fulton L, Gordon JI: **Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome.** *Cell Host & Microbe* 2008, **3:**213-223.

173. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, et al: **Linking long-term dietary patterns with gut microbial enterotypes.** *Science* 2011, **334:**105-108.

174. Candela M, Biagi E, Maccaferri S, Turroni S, Brigidi P: **Intestinal microbiota is a plastic factor responding to environmental changes.** *Trends Microbiol* 2012, **20:**385-391.

175. Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A, Dantas G, Gordon JI: **Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108:**6252-6257.

176. Antonopoulos DA, Huse SM, Morrison HG, Schmidt TM, Sogin ML, Young VB: **Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation.** *Infection and immunity* 2009, **77:**2367-2375.

177. Preidis GA, Versalovic J: **Targeting the human microbiome with antibiotics, probiotics, and prebiotics: gastroenterology enters the metagenomics era.** *Gastroenterology* 2009, **136:**2015-2031.

178. McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, Henrissat B, Oozeer R, Cools-Portier S, Gobert G, et al: **The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins.** *Science translational medicine* 2011, **3:**106ra106.

179. Diaz Heijtz R, Wang S, Anuar F, Qian Y, Bjorkholm B, Samuelsson A, Hibberd ML, Forssberg H, Pettersson S: **Normal gut microbiota modulates brain development and behavior.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108:**3047-3052.

180. Neu J: **Perinatal and neonatal manipulation of the intestinal microbiome: a note of caution.** *Nutrition reviews* 2007, **65:**282-285.

181. Borody TJ, Warren EF, Leis SM, Surace R, Ashman O, Siarakas S: **Bacteriotherapy using fecal flora: toying with human motions.** *Journal of clinical gastroenterology* 2004, **38:**475-483.

182. Seksik P, Sokol H, Lepage P, Vasquez N, Manichanh C, Mangin I, Pochart P, Dore J, Marteau P: **Review article: the role of bacteria in onset and perpetuation of inflammatory bowel disease.** *Alimentary pharmacology & therapeutics* 2006, **24 Suppl 3:**11-18.

183. Johnson S, Kent SA, O'Leary KJ, Merrigan MM, Sambol SP, Peterson LR, Gerding DN: **Fatal pseudomembranous colitis associated with a variant clostridium difficile strain not detected by toxin A immunoassay.** *Annals of internal medicine* 2001, **135:**434-438.

184. Kato N, Ou CY, Kato H, Bartley SL, Brown VK, Dowell VR, Jr., Ueno K: **Identification of toxigenic Clostridium difficile by the polymerase chain reaction.** *Journal of clinical microbiology* 1991, **29:**33-37.

185. McFarland LV, Surawicz CM, Greenberg RN, Fekety R, Elmer GW, Moyer KA, Melcher SA, Bowen KE, Cox JL, Noorani Z, et al.: **A randomized placebo-controlled trial of Saccharomyces boulardii in combination**

**with standard antibiotics for Clostridium difficile disease.** *JAMA : the journal of the American Medical Association* 1994, **271:**1913-1918.

186. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Thurber RLV, Knight R: **Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.** *Nature biotechnology* 2013, **31:**814-821.

187. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP: **Human gut microbiome viewed across age and geography.** *Nature* 2012, **486:**222-227.

188. Ursell LK, Haiser HJ, Van Treuren W, Garg N, Reddivari L, Vanamala J, Dorrestein PC, Turnbaugh PJ, Knight R: **The intestinal metabolome: an intersection between microbiota and host.** *Gastroenterology* 2014, **146:**1470-1476.

189. Tringe SG, Hugenholtz P: **A renaissance for the pioneering 16S rRNA gene.** *Curr Opin Microbiol* 2008, **11:**442-446.

190. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, et al: **Diet rapidly and reproducibly alters the human gut microbiome.** *Nature* 2013.

191. Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, Zhang M, Oh PL, Nehrenberg D, Hua K, et al: **Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors.** *Proc Natl Acad Sci U S A* 2010, **107:**18933-18938.

192. Haiser HJ, Turnbaugh PJ: **Is it time for a metagenomic basis of therapeutics?** *Science* 2012, **336:**1253-1255.

193. Maurice CF, Haiser HJ, Turnbaugh PJ: **Xenobiotics shape the physiology and gene expression of the active human gut microbiome.** *Cell* 2013, **152:**39-50.

194. Marcobal A, Kashyap PC, Nelson TA, Aronov PA, Donia MS, Spormann A, Fischbach MA, Sonnenburg JL: **A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice.** *ISME J* 2013, **7:**1933-1943.

195. Penders J, Stobberingh EE, Savelkoul PH, Wolffs PF: **The human microbiome as a reservoir of antimicrobial resistance.** *Front Microbiol* 2013, **4:**87.

196. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, et al: **The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates.** *J Bacteriol* 2008, **190:**6881-6893.

197. Cryan JF, Dinan TG: **Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour.** *Nat Rev Neurosci* 2012, **13:**701-712.

198. Dettmer K, Aronov PA, Hammock BD: **Mass spectrometry-based metabolomics.** *Mass Spectrom Rev* 2007, **26:**51-78.

199. Slupsky CM: **Nuclear magnetic resonance-based analysis of urine for the rapid etiological diagnosis of pneumonia.** *Expert Opin Med Diagn* 2011, **5:**63-73.

200. Rath CM, Alexandrov T, Higginbottom SK, Song J, Milla ME, Fischbach MA, Sonnenburg JL, Dorrestein PC: **Molecular analysis of model gut microbiotas by imaging mass spectrometry and nanodesorption electrospray ionization reveals dietary metabolite transformations.** *Anal Chem* 2012, **84:**9259-9267.

201. Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM, et al: **Mass spectral molecular networking of living microbial colonies.** *Proc Natl Acad Sci U S A* 2012, **109:**E1743-1752.

202. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, et al: **HMDB 3.0--The Human Metabolome Database in 2013.** *Nucleic Acids Res* 2013, **41:**D801-807.

203. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G: **METLIN: a metabolite mass spectral database.** *Ther Drug Monit* 2005, **27:**747-751.

204. Sud M, Fahy E, Cotter D, Brown A, Dennis EA, Glass CK, Merrill AH, Jr., Murphy RC, Raetz CR, Russell DW, Subramaniam S: **LMSD: LIPID MAPS structure database.** *Nucleic Acids Res* 2007, **35:**D527-532.

205. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, et al: **MassBank: a public repository for sharing mass spectral data for life sciences.** *J Mass Spectrom* 2010, **45:**703-714.

206. Stein SE: **Chemical substructure identification by mass spectral library searching.** *J Am Soc Mass Spectrom* 1995, **6:**644-655.

207. Yang JY, Sanchez LM, Rath CM, Liu X, Boudreau PD, Bruns N, Glukhov E, Wodtke A, de Felicio R, Fenner A, et al: **Molecular networking as a dereplication strategy.** *J Nat Prod* 2013, **76:**1686-1699.

208. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, Siuzdak G: **Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites.** *Proc Natl Acad Sci U S A* 2009, **106:**3698-3703.

209. Martin FP, Sprenger N, Yap IK, Wang Y, Bibiloni R, Rochat F, Rezzi S, Cherbut C, Kochhar S, Lindon JC, et al: **Panorganismal gut microbiome-host metabolic crosstalk.** *J Proteome Res* 2009, **8:**2090-2105.

210. Machiels K, Joossens M, Sabino J, De Preter V, Arijs I, Eeckhaut V, Ballet V, Claes K, Van Immerseel F, Verbeke K, et al: **A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis.** *Gut* 2013.

211. Wang W, Chen L, Zhou R, Wang X, Song L, Huang S, Wang G, Xia B: **Increased Proportion of Bifidobacterium and the Lactobacillus group and Loss of Butyrate-producing Bacteria in Inflammatory Bowel Disease.** *J Clin Microbiol* 2013.

212. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto JM, Kennedy S, et al: **Richness of human gut microbiome correlates with metabolic markers.** *Nature* 2013, **500:**541-546.

213. Fang S, Evans RM: **Microbiology: Wealth management in the gut.** *Nature* 2013, **500:**538-539.

214. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, et al: **Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.** *Nat Biotechnol* 2013, **31:**814-821.

215. Larsen N, Vogensen FK, van den Berg FW, Nielsen DS, Andreasen AS, Pedersen BK, Al-Soud WA, Sorensen SJ, Hansen LH, Jakobsen M: **Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults.** *PLoS ONE* 2010, **5:**e9085.

216. Heederik D, von Mutius E: **Does diversity of environmental microbial exposure matter for the occurrence of allergy and asthma?** *J Allergy Clin Immunol* 2012, **130:**44-50.

217. Blaser M: **Antibiotic overuse: Stop the killing of beneficial bacteria.** *Nature* 2011, **476:**393-394.

218. Sellitto M, Bai G, Serena G, Fricke WF, Sturgeon C, Gajer P, White JR, Koenig SS, Sakamoto J, Boothe D, et al: **Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants.** *PLoS One* 2012, **7:**e33387.

219. Sela DA, Chapman J, Adeuya A, Kim JH, Chen F, Whitehead TR, Lapidus A, Rokhsar DS, Lebrilla CB, German JB, et al: **The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome.** *Proc Natl Acad Sci U S A* 2008, **105:**18964-18969.

220. Scheline RR: **Metabolism of foreign compounds by gastrointestinal microorganisms.** *Pharmacol Rev* 1973, **25:**451-523.

221. Goldman P, Peppercorn MA, Goldin BR: **Metabolism of drugs by microorganisms in the intestine.** *Am J Clin Nutr* 1974, **27:**1348-1355.

222. Goldman P: **Biochemical pharmacology of the intestinal flora.** *Annu Rev Pharmacol Toxicol* 1978, **18:**523-539.

223. Clayton TA, Baker D, Lindon JC, Everett JR, Nicholson JK: **Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106:**14728-14733.

224. Claus SP, Ellero SL, Berger B, Krause L, Bruttin A, Molina J, Paris A, Want EJ, de Waziers I, Cloarec O, et al: **Colonization-induced host-gut microbial metabolic interaction.** *MBio* 2011, **2:**e00271-00210.

225. Sousa T, Paterson R, Moore V, Carlsson A, Abrahamsson B, Basit AW: **The gastrointestinal microbiota as a site for the biotransformation of drugs.** *Int J Pharm* 2008, **363:**1-25.

226. Haiser HJ, Turnbaugh PJ: **Developing a metagenomic view of xenobiotic metabolism.** *Pharmacol Res* 2013, **69:**21-31.

227. Saha JR, Butler VP, Jr., Neu HC, Lindenbaum J: **Digoxin-inactivating bacteria: identification in human gut flora.** *Science* 1983, **220:**325-327.

228. Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ: **Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta.** *Science* 2013, **341:**295-298.

229. Sperry JF, Wilkins TD: **Arginine, a growth-limiting factor for Eubacterium lentum.** *J Bacteriol* 1976, **127:**780-784.

230. Wallace BD, Wang H, Lane KT, Scott JE, Orans J, Koo JS, Venkatesh M, Jobin C, Yeh LA, Mani S, Redinbo MR: **Alleviating cancer drug toxicity by inhibiting a bacterial enzyme.** *Science* 2010, **330:**831-835.

231. Iida N, Dzutsev A, Stewart CA, Smith L, Bouladoux N, Weingarten RA,

Molina DA, Salcedo R, Back T, Cramer S, et al: **Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment.** *Science* 2013, **342:**967-970.

232. Viaud S, Saccheri F, Mignot G, Yamazaki T, Daillere R, Hannani D, Enot DP, Pfirschke C, Engblom C, Pittet MJ, et al: **The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide.** *Science* 2013, **342:**971-976.

233. Friedman J, Alm EJ: **Inferring correlation networks from genomic survey data.** *PLoS Comput Biol* 2012, **8:**e1002687.

234. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C: **Microbial co-occurrence relationships in the human microbiome.** *PLoS Comput Biol* 2012, **8:**e1002606.

235. Carmody RN, Weintraub GS, Wrangham RW: **Energetic consequences of thermal and nonthermal food processing.** *Proc Natl Acad Sci U S A* 2011, **108:**19199-19203.

236. Carmody RN, Wrangham RW: **The energetic significance of cooking.** *J Hum Evol* 2009, **57:**379-391.

237. Camire ME, Kubow S, Donnelly DJ: **Potatoes and human health.** *Crit Rev Food Sci Nutr* 2009, **49:**823-840.

238. **Potato world: Production and Consumption - International Year of the Potato 2008** [*http://www.fao.org/potato-2008/en/world/index.html*]

239. **Vitamin C. Fact Sheet for Health Professionals.** [*http://ods.od.nih.gov/factsheets/VitaminC-HealthProfessional/*]

240. **United States Department of Agriculture. Household USDA Foods Fact Sheet.** [*http://www.fns.usda.gov/sites/default/files/HHFS_POTATOES_RED_FRESH_Dec2012.pdf*]

241. Friedman M: **Potato glycoalkaloids and metabolites: roles in the plant**

**and in the diet.** *J Agric Food Chem* 2006, **54:**8655-8681.

242. Michalak J, Gujska E, Klepacka J: **The effect of domestic preparation of some potato products on acrylamide content.** *Plant Foods Hum Nutr* 2011, **66:**307-312.

243. Bethke PC, Bussan AJ: **Acrylamide in Processed Potato Products.** *American Journal of Potato Research* 2013, **90:**403-424.

244. Madiwale GP, Reddivari L, Stone M, Holm DG, Vanamala J: **Combined effects of storage and processing on the bioactive compounds and pro-apoptotic properties of color-fleshed potatoes in human colon cancer cells.** *J Agric Food Chem* 2012, **60:**11088-11096.

245. **Guidance for Industry: A Food Labeling Guide** [*http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/LabelingNutrition/ucm064894.htm*]

246. Dominy NJ, Vogel ER, Yeakel JD, Constantino P, Lucas PW: **Mechanical properties of plant underground storage organs and implications for dietary models of early hominins.** *Evolutionary Biology* 2008, **35:**159-175.

247. Wrangham R, Conklin-Brittain N: **'Cooking as a biological trait'.** *Comp Biochem Physiol A Mol Integr Physiol* 2003, **136:**35-46.

248. Noah L, Guillon F, Bouchet B, Buleon A, Molis C, Gratas M, Champ M: **Digestion of carbohydrate from white beans (Phaseolus vulgaris L.) in healthy humans.** *J Nutr* 1998, **128:**977-985.

249. Walter J, Ley R: **The human gut microbiome: ecology and recent evolutionary changes.** *Annu Rev Microbiol* 2011, **65:**411-429.

250. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, et al: **The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome.** *Gigascience* 2012, **1:**7.

251. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello

EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al: **QIIME allows analysis of high-throughput community sequencing data.** *Nature Methods* 2010, **7:**335-336.

252. Vazquez-Baeza Y, Pirrung M, Gonzalez A, Knight R: **EMPeror: a tool for visualizing high-throughput microbial community data.** *Gigascience* 2013, **2:**16.

253. Kulen O, Stushnoff C, Holm DG: **Effect of cold storage on total phenolics content, antioxidant activity and vitamin C level of selected potato clones.** *J Sci Food Agric* 2013, **93:**2437-2444.

# APPENDICES

**Appendix A**. Supplemental Figures and Tables for Chapter II.

| Characteristic | |
|---|---|
| Ethnicity | non-Hispanic (2), Hispanic (4) |
| Age (years, mean ± SD) | 37 ± 7.5, range 25.8-45.5 |
| Duration of HSV (years, mean ± SD) | 4.6 ± 2.3, range 2-8 |
| Number of recurrences/year (mean ± SD) | 5.3 ± 4.5, range 2-12 |

**Table S2.1.** Summary of characteristics of subjects with recurrent GHP [90].

| Sample ID | Lower Estimate (%) | Upper Estimate (%) |
|---|---|---|
| Subject35.CVL0 | 0.0069 | 0.0689 |
| Subject35.CVL1 | 0.0159 | 0.1587 |
| Subject35.CVL3 | 0.0093 | 0.0931 |
| Subject35.CVL7 | 0.0175 | 0.1752 |
| Subject35.IVR7 | 0.0054 | 0.0542 |
| Subject48.CVL0 | 0.0390 | 0.3904 |
| Subject48.CVL1 | 0.1186 | 1.1862 |
| Subject48.CVL3 | 0.0142 | 0.1419 |
| Subject48.CVL7 | 0.0531 | 0.5309 |
| Subject48.IVR7 | 0.1295 | 1.2950 |
| Subject60.CVL0 | 0.0087 | 0.0871 |
| Subject60.CVL1 | 0.0103 | 0.1031 |
| Subject60.CVL10 | 0.0048 | 0.0477 |
| Subject60.CVL3 | 0.0222 | 0.2219 |
| Subject60.CVL7 | 0.0109 | 0.1086 |
| Subject60.IVR14 | 0.0110 | 0.1098 |
| Subject70.CVL0 | 0.0247 | 0.2471 |
| Subject70.CVL1 | 0.0098 | 0.0982 |
| Subject70.CVL10 | 0.0044 | 0.0438 |
| Subject70.CVL14 | 0.0185 | 0.1852 |
| Subject70.CVL3 | 0.0128 | 0.1285 |
| Subject70.CVL7 | 0.0111 | 0.1114 |
| Subject70.IVR14 | 0.0062 | 0.0617 |
| Subject89.CVL0 | 0.0093 | 0.0926 |
| Subject89.CVL1 | 0.0049 | 0.0486 |
| Subject89.CVL10 | 0.0122 | 0.1224 |
| Subject89.CVL14 | 0.0141 | 0.1405 |
| Subject89.CVL3 | 0.0126 | 0.1262 |
| Subject89.CVL7 | 0.0225 | 0.2247 |
| Subject89.IVR14 | 0.0231 | 0.2310 |
| Subject90.CVL0 | 0.1036 | 1.0363 |
| Subject90.CVL1 | 0.0240 | 0.2405 |
| Subject90.CVL3 | 0.0517 | 0.5168 |
| Subject90.CVL7 | 0.0407 | 0.4068 |
| Subject90.IVR7 | 0.0385 | 0.3855 |

**Table S2.2.** Conditional uncovered probabilities for each individual's vaginal microbial community.

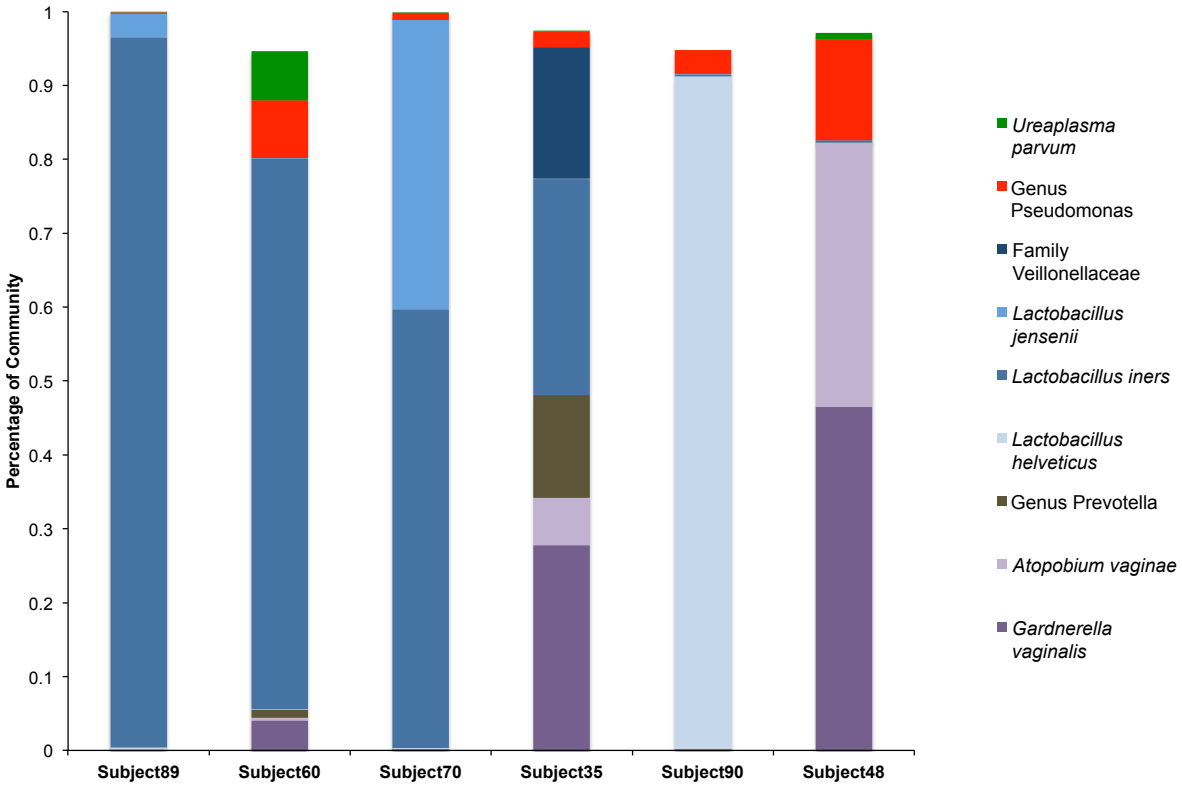


**Figure S2.1**. Taxonomic composition of the vaginal microbial communities of the 6 GHP subjects. The CVL samples from the subjects were used to assess the relative taxonomic abundance of their vaginal bacterial communities. Only taxa that comprised at least 1% of the relative abundance of any sample were graphed. Vaginal communities were dominated by Lactobacillus iners, L. helveticus, and Atopobium vaginae and Gardnerella vaginalis. The colored bars represent taxa belonging to different phyla according to the following convention: light blue to dark blue, Firmicutes; light purple to dark purple, Actinobacteria; red, Proteobacteria; brown, Bacteroidetes; green, Tenericutes.

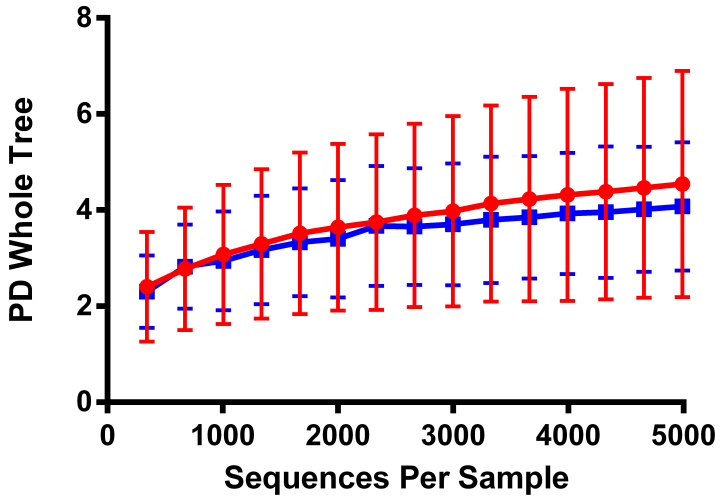


**Figure S2.2**. Comparison of α diversity in the vaginal microbial communities of GHP women (red trace, CVL samples; blue trace, IVR samples). The α diversity of the vaginal microbial communities in both groups were not significantly different using a phylogenetic α diversity metric.

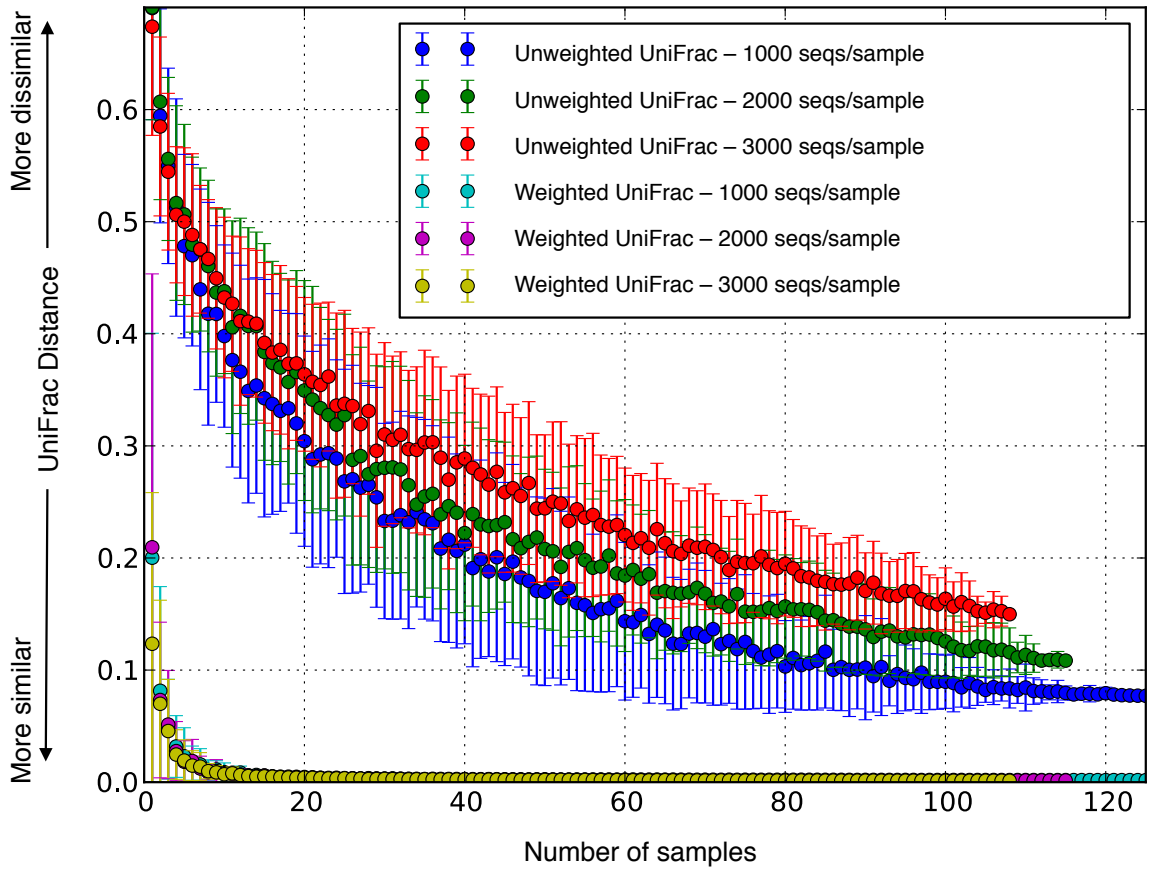


**Figure S2.3**. UniFrac distances between CVL, IVR, and HMP samples. The effect of weighted and unweighted UniFrac beta diversity distances were calculated comparing the mid-vagina samples obtained by the HMP and the GHP cohorts. Weighted UniFrac, which takes into consideration OTU abundance, was able to capture more than 99.5% of total beta diversity between the groups using just 20 samples from each group and 1000 sequences per sample. Error bars represented the standard deviation from 100 iterations of calculating the UniFrac distances from a random subsampling of the data.

# **Appendix B.** Supplemental Data associated with Chapter 4

| Taxonomy | | | Peru_Matses_mean | Venezuela_mean | Malawi_mean | USA_mean | FDR_P |
|---|---|---|---|---|---|---|---|
| Metabolism | Amino Acid Metabolism | Alanine, aspartate and glutamate metabolism [K01755] | 0.001822865 | 0.001198 | 0.0011 | 0.0011308 | 2.57E-05 |
| | | Alanine, aspartate and glutamate metabolism [K01940] | 0.001852258 | 0.001169 | 0.001 | 0.0010719 | 0.000104 |
| | | Amino acid related enzymes [K01866] | 0.002099088 | 0.001412 | 0.0015 | 0.0010285 | 3.72E-05 |
| | | Amino acid related enzymes [K01868] | 0.003432308 | 0.001626 | 0.0022 | 0.001825 | 2.52E-05 |
| | | Amino acid related enzymes [K01872] | 0.003427271 | 0.002248 | 0.0024 | 0.0020601 | 0.000364 |
| | | Amino acid related enzymes [K01874] | 0.003538653 | 0.002461 | 0.0028 | 0.0020427 | 2.14E-05 |
| | | Amino acid related enzymes [K01875] | 0.002983232 | 0.001236 | 0.0013 | 0.0011945 | 1.14E-05 |
| | | Amino acid related enzymes [K01876] | 0.003158454 | 0.00191 | 0.0026 | 0.0019364 | 0.000219 |
| | | Amino acid related enzymes [K01880] | 0.002560839 | 0.001426 | 0.001 | 0.0011146 | 6.54E-06 |
| | | Amino acid related enzymes [K01881] | 0.002553046 | 0.001358 | 0.0018 | 0.0014505 | 0.000414 |
| | | Amino acid related enzymes [K01883] | 0.002320909 | 0.001814 | 0.0014 | 0.001397 | 0.000139 |
| | | Amino acid related enzymes [K01885] | 0.002243419 | 0.002005 | 0.0017 | 0.0014286 | 0.004633 |
| | | Amino acid related enzymes [K01886] | 0.002037752 | 0.001264 | 0.0011 | 0.0012799 | 0.001961 |
| | | Amino acid related enzymes [K01887] | 0.002182758 | 0.002018 | 0.0024 | 0.0014364 | 0.000413 |
| | | Amino acid related enzymes [K01889] | 0.001737349 | 0.001225 | 0.0011 | 0.001076 | 0.001257 |
| | | Amino acid related enzymes [K01892] | 0.001724262 | 0.001156 | 0.0014 | 0.0013045 | 0.038204 |
| | | Amino acid related enzymes [K01893] | 0.002092989 | 0.000984 | 0.0009 | 0.0012947 | 3.49E-05 |
| | | Amino acid related enzymes [K04567] | 0.003048798 | 0.001747 | 0.0022 | 0.0020498 | 0.000139 |
| | | Arginine and proline metabolism [K00611] | 0.001721845 | 0.000968 | 0.0014 | 0.0010471 | 0.001749 |
| | | Arginine and proline metabolism [K01270] | 0.000927107 | 0.001292 | 0.0015 | 0.0013775 | 0.028917 |
| | | Cysteine and methionine metabolism [K00549] | 0.000771389 | 0.001255 | 0.0014 | 0.0004053 | 0.0035 |
| | | Cysteine and methionine metabolism [K01740] | 0.003642238 | 0.001961 | 0.0021 | 0.0024738 | 0.000178 |
| | | Glycine, serine and threonine metabolism [K00133] | 0.00137697 | 0.001224 | 0.0013 | 0.0009123 | 0.050802 |
| | | Glycine, serine and threonine metabolism [K01754] | 0.001162309 | 0.000997 | 0.001 | 0.0007156 | 0.049088 |
| | | Histidine metabolism [K00599] | 0.002582315 | 0.003791 | 0.0042 | 0.0038689 | 0.000966 |
| | | Histidine metabolism [K00817] | 0.000697605 | 0.000809 | 0.0011 | 0.0011574 | 0.026909 |
| | | Phenylalanine, tyrosine and tryptophan biosynthesis [K00800] | 0.000722554 | 0.001131 | 0.0013 | 0.0010482 | 0.028736 |
| | | Tryptophan metabolism [K01867] | 0.001820338 | 0.001387 | 0.0014 | 0.0010771 | 0.002996 |
| | | Tyrosine metabolism [K00680] | 0.001111452 | 0.001693 | 0.0022 | 0.001689 | 0.010797 |
| | | Valine, leucine and isoleucine biosynthesis [K00053] | 0.001441954 | 0.000766 | 0.001 | 0.0010147 | 0.002299 |
| | | Valine, leucine and isoleucine biosynthesis [K01687] | 0.002394542 | 0.001437 | 0.0018 | 0.001716 | 0.002649 |
| | | Valine, leucine and isoleucine biosynthesis [K01869] | 0.004662316 | 0.002797 | 0.0029 | 0.0021134 | 4.35E-06 |
| | | Valine, leucine and isoleucine biosynthesis [K01870] | 0.004556266 | 0.003 | 0.0032 | 0.0028623 | 4.59E-05 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Valine, leucine and isoleucine biosynthesis [K01873] | 0.004825085 | 0.002408 | 0.0023 | 0.0026124 | 2.14E-05 |
| | | Valine, leucine and isoleucine degradation [K00826] | 0.001370469 | 0.001117 | 0.0012 | 0.000929 | 0.013531 |
| | Carbohydrate Metabolism | Amino sugar and nucleotide sugar metabolism [K00820] | 0.002298518 | 0.001858 | 0.0019 | 0.0018067 | 0.047434 |
| | | Amino sugar and nucleotide sugar metabolism [K01209] | 0.001263383 | 0.002317 | 0.0027 | 0.0025134 | 0.000911 |
| | | Amino sugar and nucleotide sugar metabolism [K01443] | 0.000480795 | 0.000645 | 0.0007 | 0.0010948 | 0.000911 |
| | | Butanoate metabolism [K01652] | 0.00242262 | 0.001656 | 0.0025 | 0.0023115 | 0.033792 |
| | | C5-Branched dibasic acid metabolism [K01703] | 0.001659658 | 0.001266 | 0.0015 | 0.0010303 | 0.004497 |
| | | Citrate cycle (TCA cycle) [K01647] | 0.001641397 | 0.001211 | 0.0012 | 0.0014435 | 0.017943 |
| | | Citrate cycle (TCA cycle) [K01681] | 0.002980743 | 0.001643 | 0.0019 | 0.0015645 | 2.97E-05 |
| | | Fructose and mannose metabolism [K00100] | 0.000817646 | 0.001522 | 0.0015 | 0.0020999 | 2.52E-05 |
| | | Fructose and mannose metabolism [K02770] | 0.001125868 | 0.000688 | 0.0007 | 0.0009282 | 0.025418 |
| | | Galactose metabolism [K01190] | 0.004907241 | 0.008953 | 0.0073 | 0.0101803 | 1.13E-05 |
| | | Glycolysis / Gluconeogenesis [K00001] | 0.000933968 | 0.001415 | 0.001 | 0.0013178 | 0.003336 |
| | | Glycolysis / Gluconeogenesis [K00134] | 0.002442728 | 0.00109 | 0.0013 | 0.0009589 | 1.13E-05 |
| | | Glycolysis / Gluconeogenesis [K00845] | 0.000964382 | 0.001333 | 0.0012 | 0.0016258 | 0.035124 |
| | | Glycolysis / Gluconeogenesis [K00850] | 0.002651225 | 0.001935 | 0.0019 | 0.0028544 | 0.014005 |
| | | Glycolysis / Gluconeogenesis [K00927] | 0.001868517 | 0.00117 | 0.0016 | 0.0011178 | 0.000816 |
| | | Glycolysis / Gluconeogenesis [K01610] | 0.001925632 | 0.001333 | 0.0014 | 0.0011572 | 0.001808 |
| | | Glycolysis / Gluconeogenesis [K01624] | 0.001868812 | 0.001292 | 0.0011 | 0.0014644 | 0.018186 |
| | | Glycolysis / Gluconeogenesis [K01810] | 0.00182501 | 0.001292 | 0.0015 | 0.0012228 | 0.024969 |
| | | Glycolysis / Gluconeogenesis [K01834] | 0.003259144 | 0.002905 | 0.0028 | 0.0023875 | 0.007234 |
| | | Glycolysis / Gluconeogenesis [K04072] | 0.002566447 | 0.000753 | 0.001 | 0.0011045 | 1.13E-05 |
| | | Glyoxylate and dicarboxylate metabolism [K01091] | 0.000644695 | 0.001224 | 0.0012 | 0.0013826 | 0.000413 |
| | | Pentose and glucuronate interconversions [K00012] | 0.001853229 | 0.001428 | 0.0017 | 0.0011259 | 0.013047 |
| | | Pentose and glucuronate interconversions [K00848] | 0.000464059 | 0.001092 | 0.0006 | 0.0008988 | 0.000599 |
| | | Pentose phosphate pathway [K00615] | 0.003235503 | 0.002301 | 0.0022 | 0.0027552 | 0.006817 |
| | | Pentose phosphate pathway [K00948] | 0.001514156 | 0.001424 | 0.0014 | 0.000984 | 0.02239 |
| | | Pyruvate metabolism [K00027] | 0.001203082 | 0.000498 | 0.0009 | 0.0007088 | 0.007234 |
| | | Pyruvate metabolism [K00656] | 0.00348238 | 0.002181 | 0.0024 | 0.0028231 | 0.003883 |
| | | Pyruvate metabolism [K00925] | 0.001678319 | 0.001332 | 0.0012 | 0.0009514 | 0.003714 |
| | | Pyruvate metabolism [K01006] | 0.005620081 | 0.002594 | 0.0021 | 0.0021615 | 4.35E-06 |
| | | Pyruvate metabolism [K01571] | 0.001706196 | 0.000981 | 0.0015 | 0.0016012 | 0.031424 |
| | | Starch and sucrose metabolism [K00688] | 0.004041879 | 0.003014 | 0.0028 | 0.0038158 | 0.001856 |
| | | Starch and sucrose metabolism [K00700] | 0.003146848 | 0.00218 | 0.003 | 0.0022168 | 0.002649 |
| | | Starch and sucrose metabolism [K00702] | 0.001292308 | 0.000847 | 0.0019 | 0.0007265 | 0.007234 |
| | | Starch and sucrose metabolism [K05349] | 0.003107641 | 0.004024 | 0.005 | 0.0061795 | 3.41E-05 |
| | Energy Metabolism | Carbon fixation pathways in prokaryotes [K01938] | 0.002267101 | 0.001209 | 0.0018 | 0.0015687 | 0.000413 |
| | | Carbon fixation pathways in prokaryotes [K03737] | 0.00588303 | 0.00304 | 0.0033 | 0.0030477 | 2.39E-05 |
| | | Methane metabolism [K00532] | 0.000540847 | 0.000983 | 0.001 | 0.0008525 | 0.003099 |
| | | Methane metabolism [K00600] | 0.001767361 | 0.001129 | 0.001 | 0.0009653 | 9.56E-05 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Methane metabolism [K00831] | 0.001455574 | 0.000996 | 0.0011 | 0.0007814 | 0.000271 |
| | | Nitrogen metabolism [K00262] | 0.002265971 | 0.001305 | 0.0012 | 0.0013297 | 0.000107 |
| | | Nitrogen metabolism [K00265] | 0.00364129 | 0.002622 | 0.0027 | 0.0024046 | 0.001491 |
| | | Nitrogen metabolism [K00266] | 0.003893481 | 0.003458 | 0.003 | 0.0026134 | 0.000848 |
| | | Nitrogen metabolism [K01915] | 0.004801747 | 0.003364 | 0.0039 | 0.0034661 | 4.89E-05 |
| | | Nitrogen metabolism [K01953] | 0.002085745 | 0.001294 | 0.0012 | 0.0017961 | 0.013921 |
| | | Oxidative phosphorylation [K01507] | 0.001924618 | 0.0014 | 0.0011 | 0.0017828 | 0.033499 |
| | | Oxidative phosphorylation [K02111] | 0.002238571 | 0.002006 | 0.0012 | 0.0018867 | 0.025314 |
| | | Oxidative phosphorylation [K02112] | 0.002519486 | 0.001397 | 0.0015 | 0.0013638 | 4.40E-05 |
| | | Oxidative phosphorylation [K02117] | 0.001649233 | 0.000684 | 0.0006 | 0.00128 | 0.010503 |
| | | Sulfur metabolism [K01738] | 0.001750195 | 0.001252 | 0.0015 | 0.0011002 | 0.000599 |
| | Enzyme Families | Peptidases [K01358] | 0.001178156 | 0.000794 | 0.0011 | 0.0008581 | 0.033499 |
| | | Peptidases [K03797] | 0.001204376 | 0.002644 | 0.0024 | 0.0017403 | 0.003567 |
| | | Peptidases [K03798] | 0.004866308 | 0.002582 | 0.0035 | 0.0028721 | 1.99E-05 |
| | | Protein kinases [K07636] | 0.000710611 | 0.001102 | 0.0013 | 0.0016922 | 0.000164 |
| | | Protein kinases [K07646] | 0.000991022 | 0.00085 | 0.0007 | 0.001452 | 0.018102 |
| | | Protein kinases [K07718] | 0.000675262 | 0.000767 | 0.0009 | 0.0016147 | 0.006113 |
| | Glycan Biosynthesis and Metabolism | Glycosyltransferases [K00785] | 0.000427519 | 0.000725 | 0.0005 | 0.0006414 | 0.035744 |
| | | Glycosyltransferases [K05366] | 0.001840779 | 0.002717 | 0.003 | 0.0020209 | 0.007995 |
| | | Other glycan degradation [K01206] | 0.000924277 | 0.00152 | 0.0011 | 0.0022245 | 0.000313 |
| | | Peptidoglycan biosynthesis [K01000] | 0.001027559 | 0.001157 | 0.0013 | 0.0008668 | 0.051412 |
| | | Peptidoglycan biosynthesis [K05515] | 0.001077006 | 0.001724 | 0.0016 | 0.0015237 | 0.007234 |
| | Lipid Metabolism | Fatty acid biosynthesis [K00059] | 0.001166887 | 0.001332 | 0.0014 | 0.0016925 | 0.017517 |
| | | Fatty acid biosynthesis [K00648] | 0.000614814 | 0.00078 | 0.0007 | 0.000922 | 0.039443 |
| | | Fatty acid metabolism [K01897] | 0.001384383 | 0.00261 | 0.0024 | 0.0020436 | 0.003099 |
| | | Glycerolipid metabolism [K00864] | 0.001514497 | 0.001089 | 0.0007 | 0.0010201 | 0.017943 |
| | | Glycerophospholipid metabolism [K00995] | 0.000457825 | 0.000444 | 0.0006 | 0.0007631 | 0.048765 |
| | Metabolism of Cofactors and Vitamins | Nicotinate and nicotinamide metabolism [K00763] | 0.001669607 | 0.001059 | 0.0006 | 0.0010029 | 0.004082 |
| | | Porphyrin and chlorophyll metabolism [K02495] | 0.001216552 | 0.00129 | 0.0017 | 0.001638 | 0.037752 |
| | | Vitamin B6 metabolism [K00868] | 0.000475519 | 0.00078 | 0.0007 | 0.0005543 | 0.024145 |
| | Metabolism of Other Amino Acids | D-Alanine metabolism [K01775] | 0.000890937 | 0.001573 | 0.0011 | 0.0012096 | 0.008906 |
| | | Selenocompound metabolism [K11717] | 0.000686169 | 0.001052 | 0.0008 | 0.0005308 | 0.038716 |
| | Metabolism of Terpenoids and Polyketides | Polyketide sugar unit biosynthesis [K00973] | 0.001352799 | 0.000834 | 0.0013 | 0.0009172 | 0.009327 |
| | Nucleotide Metabolism | Purine metabolism [K00088] | 0.002379928 | 0.001371 | 0.0017 | 0.0015685 | 0.000413 |
| | | Purine metabolism [K00525] | 0.00083023 | 0.001373 | 0.0004 | 0.0014626 | 0.001664 |
| | | Purine metabolism [K00527] | 0.00535455 | 0.003037 | 0.0027 | 0.0029906 | 1.32E-05 |
| | | Purine metabolism [K00962] | 0.003455857 | 0.001978 | 0.0024 | 0.0018512 | 2.52E-05 |
| | | Purine metabolism [K01756] | 0.00206292 | 0.00121 | 0.0012 | 0.0012766 | 0.001941 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Purine metabolism [K01923] | 0.001101497 | 0.000659 | 0.0012 | 0.0007427 | 0.007517 |
| | | Purine metabolism [K01933] | 0.001413931 | 0.001077 | 0.001 | 0.0010603 | 0.042075 |
| | | Purine metabolism [K01939] | 0.002074794 | 0.001333 | 0.0015 | 0.0013174 | 0.010032 |
| | | Purine metabolism [K01951] | 0.00256111 | 0.001601 | 0.0016 | 0.0017375 | 0.002993 |
| | | Purine metabolism [K01952] | 0.004622783 | 0.00338 | 0.0034 | 0.0030495 | 0.000967 |
| | | Purine metabolism [K02337] | 0.004120104 | 0.003228 | 0.0034 | 0.0032204 | 0.003883 |
| | | Purine metabolism [K03040] | 0.001516134 | 0.001022 | 0.0014 | 0.0008283 | 0.001749 |
| | | Purine metabolism [K03043] | 0.006942539 | 0.003939 | 0.0044 | 0.0031322 | 3.76E-06 |
| | | Purine metabolism [K03046] | 0.006577827 | 0.00429 | 0.0044 | 0.0037691 | 1.56E-05 |
| | | Purine metabolism [K03763] | 0.003313575 | 0.001896 | 0.0019 | 0.0020581 | 0.001549 |
| | | Pyrimidine metabolism [K00609] | 0.001382088 | 0.00101 | 0.001 | 0.0008737 | 0.011569 |
| | | Pyrimidine metabolism [K00876] | 0.001204989 | 0.001858 | 0.0017 | 0.0012517 | 0.018102 |
| | | Pyrimidine metabolism [K00945] | 0.000806694 | 0.001547 | 0.0016 | 0.0016964 | 3.72E-05 |
| | | Pyrimidine metabolism [K01937] | 0.002846673 | 0.001654 | 0.0019 | 0.0012703 | 2.39E-05 |
| | | Pyrimidine metabolism [K01955] | 0.00589094 | 0.003753 | 0.0036 | 0.004014 | 2.20E-05 |
| | | Pyrimidine metabolism [K01956] | 0.001293208 | 0.001361 | 0.0007 | 0.0009647 | 0.018236 |
| Environmental Information Processing | Membrane Transport | Bacterial secretion system [K03205] | 0.005594136 | 0.002848 | 0.0029 | 0.005342 | 0.004225 |
| | | Secretion system [K02238] | 0.000493764 | 0.000793 | 0.0009 | 0.0013205 | 0.00013 |
| | | Transporters [K01990] | 0.002425131 | 0.002704 | 0.003 | 0.0033267 | 0.009303 |
| | | Transporters [K01992] | 0.000770799 | 0.001494 | 0.0018 | 0.00264 | 5.47E-06 |
| | | Transporters [K02003] | 0.004752995 | 0.003525 | 0.0036 | 0.0054236 | 0.004093 |
| | | Transporters [K02004] | 0.005157821 | 0.007024 | 0.0061 | 0.0110482 | 3.76E-06 |
| | | Transporters [K02013] | 0.000529301 | 0.001255 | 0.0008 | 0.0013054 | 3.24E-05 |
| | | Transporters [K02015] | 0.000999915 | 0.001698 | 0.0013 | 0.0017355 | 0.006401 |
| | | Transporters [K02016] | 0.000891387 | 0.001132 | 0.0011 | 0.0015401 | 0.015504 |
| | | Transporters [K02025] | 0.002657483 | 0.002283 | 0.0032 | 0.0038779 | 0.050216 |
| | | Transporters [K02026] | 0.002159505 | 0.002122 | 0.0028 | 0.0038126 | 0.011612 |
| | | Transporters [K02028] | 0.002167612 | 0.001467 | 0.0016 | 0.0011219 | 0.001502 |
| | | Transporters [K02036] | 0.001148428 | 0.000579 | 0.0006 | 0.0007636 | 0.043353 |
| | | Transporters [K06147] | 0.014153341 | 0.010862 | 0.0121 | 0.0121234 | 0.024348 |
| | | Transporters [K10112] | 0.001165354 | 0.000551 | 0.0005 | 0.0007086 | 0.008668 |
| | Signal Transduction | Two-component system [K03406] | 0.000629617 | 0.000877 | 0.0019 | 0.0013546 | 0.024708 |
| Genetic Information Processing | Folding, Sorting and Degradation | Chaperones and folding catalysts [K03544] | 0.001924265 | 0.001158 | 0.0018 | 0.0014728 | 0.002598 |
| | | Chaperones and folding catalysts [K03671] | 0.000469941 | 0.000847 | 0.0009 | 0.000588 | 0.052115 |
| | | Chaperones and folding catalysts [K03695] | 0.004092958 | 0.002461 | 0.0029 | 0.0021703 | 1.13E-05 |
| | | Chaperones and folding catalysts [K04043] | 0.00476685 | 0.002151 | 0.0019 | 0.0018106 | 3.76E-06 |
| | | Chaperones and folding catalysts [K04077] | 0.002542328 | 0.001426 | 0.0018 | 0.0014033 | 4.59E-05 |
| | | Chaperones and folding catalysts [K04079] | 0.001984128 | 0.001629 | 0.0013 | 0.0013663 | 0.015504 |
| | | Protein export [K03070] | 0.004662654 | 0.002972 | 0.0034 | 0.0026506 | 3.72E-05 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Protein export [K03106] | 0.001984818 | 0.001561 | 0.0011 | 0.0010237 | 3.72E-05 |
| | | Protein export [K12257] | 0.001338677 | 0.002152 | 0.0016 | 0.0016723 | 0.050802 |
| | | RNA degradation [K03654] | 0.001897301 | 0.003043 | 0.0029 | 0.0025513 | 0.007517 |
| | | RNA degradation [K12574] | 0.001502561 | 0.000818 | 0.0012 | 0.0009347 | 0.003567 |
| | Replication and Repair | Base excision repair [K01142] | 0.001167465 | 0.000618 | 0.0006 | 0.0005072 | 0.000177 |
| | | Chromosome [K03495] | 0.002994535 | 0.002124 | 0.002 | 0.0019437 | 0.001981 |
| | | Chromosome [K03529] | 0.002195347 | 0.001571 | 0.0015 | 0.0015311 | 0.03899 |
| | | DNA repair and recombination proteins [K02346] | 0.000661711 | 0.000552 | 0.0011 | 0.0011 | 0.003099 |
| | | DNA repair and recombination proteins [K03546] | 0.000776651 | 0.001065 | 0.0011 | 0.0012514 | 0.046337 |
| | | DNA repair and recombination proteins [K03631] | 0.000897841 | 0.001268 | 0.0014 | 0.0012797 | 0.031528 |
| | | DNA replication [K03111] | 0.000554124 | 0.000998 | 0.001 | 0.0008676 | 0.006113 |
| | | DNA replication proteins [K02315] | 0.001014987 | 0.000927 | 0.0009 | 0.0018003 | 0.009295 |
| | | DNA replication proteins [K02469] | 0.00545321 | 0.003814 | 0.0035 | 0.003862 | 0.000141 |
| | | DNA replication proteins [K02470] | 0.005390871 | 0.003565 | 0.0035 | 0.0031007 | 2.14E-05 |
| | | DNA replication proteins [K03168] | 0.002657847 | 0.002085 | 0.0028 | 0.00167 | 0.000177 |
| | | DNA replication proteins [K03169] | 0.004947981 | 0.003472 | 0.0043 | 0.0057067 | 0.019785 |
| | | DNA replication proteins [K03530] | 0.00037924 | 0.000673 | 0.0006 | 0.0006427 | 0.004658 |
| | | Homologous recombination [K03551] | 0.001622305 | 0.00125 | 0.0014 | 0.0008806 | 0.000271 |
| | | Homologous recombination [K03553] | 0.002072258 | 0.000969 | 0.0008 | 0.0010364 | 4.52E-05 |
| | | Homologous recombination [K03581] | 0.001148466 | 0.001894 | 0.0017 | 0.0016888 | 0.011857 |
| | | Homologous recombination [K03655] | 0.003619475 | 0.005127 | 0.0055 | 0.0032964 | 0.001491 |
| | | Mismatch repair [K03555] | 0.002993101 | 0.002794 | 0.0028 | 0.0022241 | 0.017063 |
| | | Nucleotide excision repair [K03701] | 0.006580007 | 0.004386 | 0.0038 | 0.0037454 | 2.14E-05 |
| | | Nucleotide excision repair [K03702] | 0.003414319 | 0.001976 | 0.0021 | 0.0018399 | 1.56E-05 |
| | | Nucleotide excision repair [K03723] | 0.003995095 | 0.003347 | 0.0034 | 0.0030063 | 0.01887 |
| | Transcription | Transcription factors [K02529] | 0.001735638 | 0.003109 | 0.0033 | 0.003392 | 1.99E-05 |
| | | Transcription factors [K03497] | 0.002567805 | 0.001805 | 0.0023 | 0.0032279 | 0.002545 |
| | | Transcription factors [K03711] | 0.000438479 | 0.000526 | 0.0006 | 0.0007912 | 0.009137 |
| | | Transcription machinery [K02600] | 0.001318567 | 0.001481 | 0.0012 | 0.0009835 | 0.01049 |
| | | Transcription machinery [K03086] | 0.002118284 | 0.002017 | 0.0019 | 0.0011925 | 0.001788 |
| | | Transcription machinery [K03088] | 0.001906689 | 0.003311 | 0.0033 | 0.0049126 | 2.52E-05 |
| | Translation | Ribosome [K02886] | 0.001784963 | 0.001075 | 0.0011 | 0.0008338 | 0.000334 |
| | | Ribosome [K02950] | 0.000955603 | 0.000417 | 0.0007 | 0.0005068 | 0.000413 |
| | | Ribosome [K02982] | 0.00129874 | 0.000984 | 0.0008 | 0.0007404 | 0.002964 |
| | | Ribosome Biogenesis [K03595] | 0.000969515 | 0.000566 | 0.0006 | 0.0007813 | 0.023488 |
| | | Ribosome Biogenesis [K03977] | 0.001821278 | 0.001533 | 0.0016 | 0.001218 | 0.039655 |
| | | Ribosome Biogenesis [K03979] | 0.001774409 | 0.001102 | 0.0014 | 0.0011639 | 0.001153 |
| | | Ribosome Biogenesis [K06969] | 0.001766256 | 0.00117 | 0.0013 | 0.001256 | 0.037752 |
| | | Translation factors [K02355] | 0.00644639 | 0.004278 | 0.0043 | 0.0040333 | 2.14E-05 |
| | | Translation factors [K02356] | 0.000967717 | 0.000644 | 0.0008 | 0.0005169 | 0.005042 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Translation factors [K02358] | 0.00318607 | 0.001599 | 0.0017 | 0.0010304 | 5.47E-06 |
| | Translation factors [K02519] | 0.003302323 | 0.003067 | 0.003 | 0.0020792 | 0.001856 |
| | Translation factors [K02835] | 0.001608878 | 0.001089 | 0.0011 | 0.0009051 | 0.007565 |
| | Translation factors [K02836] | 0.001952555 | 0.001184 | 0.0015 | 0.000976 | 7.11E-05 |
| | Translation factors [K02837] | 0.001600108 | 0.001264 | 0.0009 | 0.0012566 | 0.028736 |

**Table S4.1. Matses gut microbiota are enriched for metabolic functions compared to Venezuela, Malawi, and USA.** We filtered KOs to only those shared by every sample from all populations. We ran Kruskal-Wallis to determine if there were significant differences in the abundance of KOs between countries. KOs with an FDR corrected p-value < 0.05 are reported along with the mean abundance of the KO in each country. KOs for which the Matses population in Peru was enriched compared to other countries are highlighted in blue.

| Taxonomy | | | Peru_Matses_mean | Non-Western_mean | FDR_P |
|---|---|---|---|---|---|
| Metabolism | Amino Acid Metabolism | Alanine, aspartate and glutamate metabolism [K01755] | 0.00182287 | 0.001181023 | 0.000262 |
| | | Alanine, aspartate and glutamate metabolism [K01940] | 0.00185226 | 0.00112172 | 0.000651 |
| | | Amino acid related enzymes [K01866] | 0.00209909 | 0.001448794 | 0.005323 |
| | | Amino acid related enzymes [K01868] | 0.00343231 | 0.001783154 | 0.000367 |
| | | Amino acid related enzymes [K01872] | 0.00342727 | 0.002302587 | 0.002048 |
| | | Amino acid related enzymes [K01874] | 0.00353865 | 0.002561027 | 0.002681 |
| | | Amino acid related enzymes [K01875] | 0.00298323 | 0.001247634 | 0.000215 |
| | | Amino acid related enzymes [K01876] | 0.00315845 | 0.002118734 | 0.003989 |
| | | Amino acid related enzymes [K01880] | 0.00256084 | 0.001316839 | 0.000215 |
| | | Amino acid related enzymes [K01881] | 0.00255305 | 0.001476896 | 0.003285 |
| | | Amino acid related enzymes [K01883] | 0.00232091 | 0.001706595 | 0.009838 |
| | | Amino acid related enzymes [K01886] | 0.00203775 | 0.001228585 | 0.003106 |
| | | Amino acid related enzymes [K01889] | 0.00173735 | 0.001199268 | 0.010484 |
| | | Amino acid related enzymes [K01892] | 0.00172426 | 0.0012171 | 0.021364 |
| | | Amino acid related enzymes [K01893] | 0.00209299 | 0.000970954 | 0.000215 |
| | | Amino acid related enzymes [K04567] | 0.0030488 | 0.001868614 | 0.000262 |
| | | Arginine and proline metabolism [K00611] | 0.00172185 | 0.001102801 | 0.007761 |
| | | Arginine and proline metabolism [K01270] | 0.00092711 | 0.001343203 | 0.02494 |
| | | Cysteine and methionine metabolism [K01740] | 0.00364224 | 0.001995103 | 0.000292 |
| | | Glycine, serine and threonine metabolism [K06001] | 0.00095623 | 0.000479868 | 0.029056 |
| | | Histidine metabolism [K00599] | 0.00258232 | 0.003921378 | 0.006508 |
| | | Histidine metabolism [K00765] | 0.0004507 | 0.000642679 | 0.031112 |
| | | Phenylalanine, tyrosine and tryptophan biosynthesis [K00800] | 0.00072255 | 0.001190068 | 0.023271 |
| | | Tyrosine metabolism [K00680] | 0.00111145 | 0.001842384 | 0.021364 |
| | | Valine, leucine and isoleucine biosynthesis [K00053] | 0.00144195 | 0.000825652 | 0.00256 |
| | | Valine, leucine and isoleucine biosynthesis [K01687] | 0.00239454 | 0.001533355 | 0.003106 |
| | | Valine, leucine and isoleucine biosynthesis [K01869] | 0.00466232 | 0.002829793 | 0.000282 |
| | | Valine, leucine and isoleucine biosynthesis [K01870] | 0.00455627 | 0.003061187 | 0.000262 |
| | | Valine, leucine and isoleucine biosynthesis [K01873] | 0.00482509 | 0.002388326 | 0.000262 |
| | Carbohydrate Metabolism | Amino sugar and nucleotide sugar metabolism [K00820] | 0.00229852 | 0.001863184 | 0.02713 |
| | | Amino sugar and nucleotide sugar metabolism [K01209] | 0.00126338 | 0.0024319 | 0.007228 |
| | | Butanoate metabolism [K01652] | 0.00242262 | 0.001890789 | 0.03589 |

| | | | | | |
|---|---|---|---|---|---|
| | | Citrate cycle (TCA cycle) [K01647] | 0.0016414 | 0.001198934 | 0.003989 |
| | | Citrate cycle (TCA cycle) [K01681] | 0.00298074 | 0.001710268 | 0.000367 |
| | | Fructose and mannose metabolism [K00100] | 0.00081765 | 0.001508637 | 0.002681 |
| | | Fructose and mannose metabolism [K00847] | 0.00096281 | 0.001391826 | 0.041595 |
| | | Fructose and mannose metabolism [K01818] | 0.00110025 | 0.000576577 | 0.031112 |
| | | Fructose and mannose metabolism [K02770] | 0.00112587 | 0.00068194 | 0.00434 |
| | | Galactose metabolism [K01190] | 0.00490724 | 0.008475879 | 0.000599 |
| | | Glycolysis / Gluconeogenesis [K00001] | 0.00093397 | 0.001288661 | 0.02494 |
| | | Glycolysis / Gluconeogenesis [K00134] | 0.00244273 | 0.001162466 | 0.000277 |
| | | Glycolysis / Gluconeogenesis [K00850] | 0.00265123 | 0.001925357 | 0.02713 |
| | | Glycolysis / Gluconeogenesis [K00927] | 0.00186852 | 0.001284356 | 0.010484 |
| | | Glycolysis / Gluconeogenesis [K01610] | 0.00192563 | 0.001363339 | 0.008428 |
| | | Glycolysis / Gluconeogenesis [K01624] | 0.00186881 | 0.00122896 | 0.005922 |
| | | Glycolysis / Gluconeogenesis [K04072] | 0.00256645 | 0.000823129 | 0.000203 |
| | | Glyoxylate and dicarboxylate metabolism [K01091] | 0.0006447 | 0.001218732 | 0.003989 |
| | | Pentose and glucuronate interconversions [K00848] | 0.00046406 | 0.00096132 | 0.002681 |
| | | Pentose phosphate pathway [K00615] | 0.0032355 | 0.002267194 | 0.001762 |
| | | Pyruvate metabolism [K00027] | 0.00120308 | 0.00061342 | 0.011595 |
| | | Pyruvate metabolism [K00048] | 0.00099158 | 0.000614083 | 0.009151 |
| | | Pyruvate metabolism [K00656] | 0.00348238 | 0.002254883 | 0.000974 |
| | | Pyruvate metabolism [K01006] | 0.00562008 | 0.002446809 | 0.000203 |
| | | Pyruvate metabolism [K01571] | 0.0017062 | 0.001131928 | 0.016313 |
| | | Starch and sucrose metabolism [K00688] | 0.00404188 | 0.002945595 | 0.001194 |
| | | Starch and sucrose metabolism [K00700] | 0.00314685 | 0.00241836 | 0.02494 |
| | | Starch and sucrose metabolism [K05349] | 0.00310764 | 0.004310105 | 0.013649 |
| | Energy Metabolism | Carbon fixation pathways in prokaryotes [K01938] | 0.0022671 | 0.001379527 | 0.00128 |
| | | Carbon fixation pathways in prokaryotes [K03737] | 0.00588303 | 0.003127188 | 0.000262 |
| | | Methane metabolism [K00532] | 0.00054085 | 0.000987731 | 0.003285 |
| | | Methane metabolism [K00600] | 0.00176736 | 0.001083688 | 0.001066 |
| | | Methane metabolism [K00831] | 0.00145557 | 0.001035986 | 0.013649 |
| | | Nitrogen metabolism [K00262] | 0.00226597 | 0.001286656 | 0.000277 |
| | | Nitrogen metabolism [K00265] | 0.00364129 | 0.002637779 | 0.015056 |
| | | Nitrogen metabolism [K01915] | 0.00480175 | 0.003515175 | 0.001898 |
| | | Nitrogen metabolism [K01953] | 0.00208575 | 0.001258976 | 0.002993 |

| | | | | | |
|---|---|---|---|---|---|
| | | Oxidative phosphorylation [K01507] | 0.00192462 | 0.001325915 | 0.010484 |
| | | Oxidative phosphorylation [K02111] | 0.00223857 | 0.001776885 | 0.03589 |
| | | Oxidative phosphorylation [K02112] | 0.00251949 | 0.0014185 | 0.000282 |
| | | Oxidative phosphorylation [K02117] | 0.00164923 | 0.00066083 | 0.003106 |
| | | Sulfur metabolism [K00640] | 0.00070878 | 0.001027874 | 0.021364 |
| | | Sulfur metabolism [K01738] | 0.0017502 | 0.001324658 | 0.016313 |
| | Enzyme Families | Peptidases [K01358] | 0.00117816 | 0.000883014 | 0.038909 |
| | | Peptidases [K03797] | 0.00120438 | 0.002569555 | 0.00128 |
| | | Peptidases [K03798] | 0.00486631 | 0.002846636 | 0.000552 |
| | | Protein kinases [K07636] | 0.00071061 | 0.001161365 | 0.02494 |
| | Glycan Biosynthesis and Metabolism | Glycosyltransferases [K00785] | 0.00042752 | 0.000651927 | 0.05262 |
| | | Glycosyltransferases [K05366] | 0.00184078 | 0.002811254 | 0.003285 |
| | | Other glycan degradation [K01206] | 0.00092428 | 0.00139213 | 0.048674 |
| | | Peptidoglycan biosynthesis [K05515] | 0.00107701 | 0.001691013 | 0.006508 |
| | Lipid Metabolism | Fatty acid metabolism [K01897] | 0.00138438 | 0.002556185 | 0.001762 |
| | | Glycerolipid metabolism [K00864] | 0.0015145 | 0.000967855 | 0.017814 |
| | Metabolism of Cofactors and Vitamins | Biotin metabolism [K03524] | 0.00074082 | 0.001016574 | 0.03354 |
| | | Nicotinate and nicotinamide metabolism [K00763] | 0.00166961 | 0.000938068 | 0.005323 |
| | | Thiamine metabolism [K04487] | 0.00188839 | 0.001437596 | 0.048674 |
| | | Vitamin B6 metabolism [K00868] | 0.00047552 | 0.000768325 | 0.012471 |
| | Metabolism of Other Amino Acids | D-Alanine metabolism [K01775] | 0.00089094 | 0.001449506 | 0.007761 |
| | Metabolism of Terpenoids and Polyketides | Polyketide sugar unit biosynthesis [K00973] | 0.0013528 | 0.00096776 | 0.03589 |
| | Nucleotide Metabolism | Purine metabolism [K00088] | 0.00237993 | 0.001466943 | 0.00128 |
| | | Purine metabolism [K00527] | 0.00535455 | 0.002931853 | 0.000227 |
| | | Purine metabolism [K00602] | 0.00151518 | 0.001124554 | 0.05262 |
| | | Purine metabolism [K00962] | 0.00345586 | 0.00210977 | 0.001066 |
| | | Purine metabolism [K01756] | 0.00206292 | 0.001219477 | 0.003989 |
| | | Purine metabolism [K01933] | 0.00141393 | 0.001065581 | 0.017814 |
| | | Purine metabolism [K01939] | 0.00207479 | 0.001373952 | 0.009838 |
| | | Purine metabolism [K01951] | 0.00256111 | 0.001600577 | 0.003106 |
| | | Purine metabolism [K01952] | 0.00462278 | 0.003389246 | 0.003285 |
| | | Purine metabolism [K02337] | 0.0041201 | 0.003273085 | 0.009838 |
| | | Purine metabolism [K03043] | 0.00694254 | 0.004077235 | 0.000213 |

| | | | | | |
|---|---|---|---|---|---|
| | | Purine metabolism [K03046] | 0.00657783 | 0.004324846 | 0.000915 |
| | | Purine metabolism [K03763] | 0.00331358 | 0.001907313 | 0.001898 |
| | | Pyrimidine metabolism [K00609] | 0.00138209 | 0.00100821 | 0.041595 |
| | | Pyrimidine metabolism [K00876] | 0.00120499 | 0.00180484 | 0.009151 |
| | | Pyrimidine metabolism [K00945] | 0.00080669 | 0.001563761 | 0.000292 |
| | | Pyrimidine metabolism [K01937] | 0.00284667 | 0.001726203 | 0.000974 |
| | | Pyrimidine metabolism [K01955] | 0.00589094 | 0.003706788 | 0.000215 |
| Environmental Information Processing | Membrane Transport | Bacterial secretion system [K03205] | 0.00559414 | 0.002855123 | 0.001898 |
| | | Secretion system [K02238] | 0.00049376 | 0.000816528 | 0.009151 |
| | | Transporters [K01992] | 0.0007708 | 0.001583756 | 0.000729 |
| | | Transporters [K01995] | 0.00085972 | 0.000488071 | 0.041595 |
| | | Transporters [K02003] | 0.004753 | 0.00356 | 0.010484 |
| | | Transporters [K02004] | 0.00515782 | 0.006765727 | 0.002681 |
| | | Transporters [K02013] | 0.0005293 | 0.00111622 | 0.000915 |
| | | Transporters [K02015] | 0.00099992 | 0.001585785 | 0.008428 |
| | | Transporters [K02028] | 0.00216761 | 0.001515506 | 0.045025 |
| | | Transporters [K02032] | 0.00193471 | 0.001468723 | 0.029056 |
| | | Transporters [K02036] | 0.00114843 | 0.00057659 | 0.009838 |
| | | Transporters [K06147] | 0.01415334 | 0.011221717 | 0.012471 |
| | | Transporters [K10112] | 0.00116535 | 0.000546761 | 0.001762 |
| | Signal Transduction | Two-component system [K03406] | 0.00062962 | 0.001180852 | 0.031112 |
| Genetic Information Processing | Folding, Sorting and Degradation | Chaperones and folding catalysts [K03544] | 0.00192427 | 0.00135357 | 0.010484 |
| | | Chaperones and folding catalysts [K03671] | 0.00046994 | 0.000862848 | 0.012471 |
| | | Chaperones and folding catalysts [K03695] | 0.00409296 | 0.002580962 | 0.000292 |
| | | Chaperones and folding catalysts [K04043] | 0.00476685 | 0.002071802 | 0.000203 |
| | | Chaperones and folding catalysts [K04077] | 0.00254233 | 0.001526878 | 0.000974 |
| | | Chaperones and folding catalysts [K04079] | 0.00198413 | 0.001546267 | 0.041595 |
| | | Protein export [K03070] | 0.00466265 | 0.003089494 | 0.001434 |
| | | Protein export [K03106] | 0.00198482 | 0.001439827 | 0.008428 |
| | | Protein export [K12257] | 0.00133868 | 0.002007311 | 0.029056 |
| | | RNA degradation [K03654] | 0.0018973 | 0.002998105 | 0.002048 |
| | | RNA degradation [K12574] | 0.00150256 | 0.000936687 | 0.012471 |
| | Replication and Repair | Base excision repair [K01142] | 0.00116747 | 0.000623263 | 0.003285 |
| | | Chromosome [K03495] | 0.00299454 | 0.002081906 | 0.003106 |

| | | | | | |
|---|---|---|---|---|---|
| | | Chromosome [K03529] | 0.00219535 | 0.001550933 | 0.03354 |
| | | DNA repair and recombination proteins [K03631] | 0.00089784 | 0.001306848 | 0.029056 |
| | | DNA replication [K03111] | 0.00055412 | 0.00101059 | 0.005323 |
| | | DNA replication proteins [K02469] | 0.00545321 | 0.003711676 | 0.000651 |
| | | DNA replication proteins [K02470] | 0.00539087 | 0.003558727 | 0.000282 |
| | | DNA replication proteins [K03169] | 0.00494798 | 0.003713382 | 0.03354 |
| | | DNA replication proteins [K03530] | 0.00037924 | 0.000652653 | 0.007761 |
| | | Homologous recombination [K03553] | 0.00207226 | 0.000922351 | 0.000262 |
| | | Homologous recombination [K03581] | 0.00114847 | 0.001830464 | 0.002681 |
| | | Homologous recombination [K03655] | 0.00361948 | 0.005241007 | 0.00256 |
| | | Nucleotide excision repair [K03701] | 0.00658001 | 0.004222094 | 0.000599 |
| | | Nucleotide excision repair [K03702] | 0.00341432 | 0.002023284 | 0.000262 |
| | Transcription | Transcription factors [K02529] | 0.00173564 | 0.003167504 | 0.000215 |
| | | Transcription factors [K03497] | 0.00256781 | 0.001950244 | 0.038909 |
| | | Transcription machinery [K03088] | 0.00190669 | 0.003296141 | 0.00434 |
| | Translation | Ribosome [K02886] | 0.00178496 | 0.0010839 | 0.023271 |
| | | Ribosome [K02950] | 0.0009556 | 0.000498859 | 0.002048 |
| | | Ribosome [K02967] | 0.00115981 | 0.000816373 | 0.03589 |
| | | Ribosome [K02982] | 0.00129874 | 0.000940986 | 0.045025 |
| | | Ribosome Biogenesis [K03595] | 0.00096952 | 0.000585866 | 0.003106 |
| | | Ribosome Biogenesis [K03979] | 0.00177441 | 0.001188236 | 0.00434 |
| | | Ribosome Biogenesis [K06969] | 0.00176626 | 0.001209224 | 0.016313 |
| | | Translation factors [K02355] | 0.00644639 | 0.004281995 | 0.000262 |
| | | Translation factors [K02358] | 0.00318607 | 0.001629769 | 0.000262 |
| | | Translation factors [K02835] | 0.00160888 | 0.00109218 | 0.031112 |
| | | Translation factors [K02836] | 0.00195256 | 0.001285814 | 0.007761 |

**Table S4.2. Matses gut microbiota are enriched for metabolic functions compared to Venezuela and Malawi.** We filtered KOs to only those shared by every sample from all populations. We ran Kruskal-Wallis to determine if there were significant differences in the abundance of KOs between countries. KOs with an FDR corrected p-value < 0.05 are reported along with the mean abundance of the KO in each country. KOs for which the Matses population in Peru was enriched compared to other countries are highlighted in blue.

**Appendix B** – Supplemental Methods and Data for Chapter VI

**Methods**

Methanol was supplied by EMD chemicals (Philadelphia, MA). Monobasic ammonium phosphate was procured from Avantor Performance Materials (Phillipsburg, NJ). Carrez I and II solutions, acetonitrile, acetone and chloroform were purchased from Fisher Scientific (Fenton, MO). α –solanine and α-chaconine standards were obtained from Indofine (NJ, USA). Acrylamide, chlorogenic acid, caffeic acid, trigonelline and ascorbic acid (vitamin C) standards were obtained from Sigma-Aldrich (St.Louis, MO). Cyanidin and petunidin-3-glucoside were procured from Indofine (Hillsborough, NJ). Dithiothreitol, metaphosphoric acid, sodium formate and formic acid were purchased from VWR International (Radnor, PA).

*Potato cultivar and processing methods.* Potatoes of six commercial cultivars (Atlantic & Russet Burbank, white-fleshed; All Red & Mountain Rose, red-fleshed; All Blue & Purple Majesty, purple-fleshed) were grown at the San Luis Valley Research Center, Colorado State University, Center, CO, USA. Potatoes of each cultivar were then processed via six different methods (baked, chipped, fried, microwaved, raw and steamed). All potatoes were washed before processing. Raw samples from each cultivar were diced (with skin) into pieces weighing 7 ±1g and stored at -20$^{o}$C until analysis. Medium sized potatoes (6 to 7 oz) each wrapped in food-grade aluminum foil and pierced approximately 1.5 cm deep at 3 cm intervals were baked for 1 hour in a conventional oven preheated to 204$^{o}$C (400$^{o}$F), then

allowed to cool for 30 minutes. After cooling, potatoes were diced (with skin) and stored at -20°C until analysis. Chip slices were made using a Ditto Dean Food Prep industrial slicer (TR23) with a C-2 blade, and fry strips were cut with a C10 and FS-10 blade. Raw chips and fries were washed under running warm water for approximately 1 minute to remove any water-soluble sugars present on the surface and placed in strainer trays to remove excess water. Chips and fries were fried in a 5-liter capacity fryer (APW Wyott EF-30-208-2) with Bakers & Chefs Clear Frying Oil (Sam Club, Bentonville, AR). Chips were fried at 185°C for 2 minutes. Fried potatoes were French fried at 190°C for 3.5 minutes. After frying, potatoes were placed on paper towels to absorb any excess oil, allowed to cool for 10-15 minutes, and stored at -20°C until analysis. For microwaving, medium sized potatoes were pierced twice with a fork on each side, then cooked at the outer edge of a 1,100 watt microwave oven for 3.5 min. Potatoes were flipped over after 1.5 minutes. For steaming, a sieved double-boiler was filled halfway with water. After water reached boiling, the temperature was reduced to medium. A steamer with potatoes was placed in the bottom pan, covered with a lid and cooked for 30 minutes. After cooling, potatoes were diced (with skin) and stored at -20°C until analysis.

*Moisture Content Analysis.* Moisture content of potato samples was analyzed using a laboratory scale infrared moisture content analyzer (OHAUS MB 45). Baked, microwaved, steamed, and raw samples were analyzed with a step profile from 200°C to 105°C. Chip and fry samples were analyzed with a fast profile at 95°C and

110°C, respectively, due to low moisture content and to avoid charring. Three randomized samples were analyzed for each treatment group.

*Preparation of Extracts for Metabolomics Analysis.* Potato samples were extracted for metabolomics analysis according to published protocols [244] with minor modifications. Samples were homogenized with acidified methanol (80%, with 0.1% v/v formic acid). Homogenates were poured into chloroform resistant tubes and vortexed every 15 minutes for 1 hour. Chloroform was added to the tubes to separate the lipids, and the tubes were vortexed every 10 minutes for 30 minutes. The tubes were centrifuged at 3220 g for 10 minutes and stored overnight at 4°C. Millipore water (5 mL) was carefully added to the top layer of chip samples before overnight storage at 4°C to aid separation. Methanolic phase was collected and stored at -20°C for further analysis.

*Untargeted Analysis using UPLC-Q-TOF-MS.* Injections (1 uL) were performed on a Waters Acquity UPLC system. Separation was performed using a Waters Acquity UPLC T3 column (1.8 μM, 1.0 x 100 mm), using a gradient from solvent A (water, 0.1% formic acid) to solvent B (Acetonitrile, 0.1% formic acid). Injections were made in 100% A, which was held for one minute; a 12 minute linear gradient to 95% B was applied, and held at 95% B for 3 minutes, returned to starting conditions over 0.05 minutes, and allowed to re-equilibrate for 3.95 minutes. Flow rate was constant at 200 μL/min for the duration of the run. The column was held

at 50°C, samples were held at 5°C. Column eluent was infused into a Waters Xevo G2 Q-TOF MS fitted with an electrospray source. Data was collected in positive ion mode, scanning from 50-1200 at a rate of 0.2 seconds per scan, alternating between MS and $MS^E$ mode. Collision energy was set to 6 V for MS mode, and ramped from 15-30 V for $MS^E$ mode. Calibration was performed prior to sample analysis via infusion of sodium formate solution, with mass accuracy within 1 ppm. The capillary voltage was held at 2200 V, the source temp at 150°C, and the desolvation temperature at 350°C at a nitrogen desolvation gas flow rate of 800 L/hr. Metabolites were identified via spectral matching with MassBank [205] and Metlin [203] databases and/or targeted metabolomics using pooled samples and standards.

*Processing and analysis of metabolites.* Intensities were transformed using a log (x+1) conversion to account for zeros. Bray Curtis distances were calculated on BIOM tables[250] of metabolites and visualized using principal coordinates analysis in QIIME [251] and Emperor[252]. ANOSIM and PERMANOVA were performed using PRIMER v6 (Clarke, KR, Gorley, RN, 2006. PRIMER v6: User Manual/Tutorial) on the distance matrices (Bray-Curtis distances) for the untargeted metabolites and for the identified vitamins and minerals; p-values were corrected using False Discovery Rate. Two-way ANOVA was performed on the full table of unidentified metabolites and on targeted metabolites in R. For the untargeted metabolomics, the log transformed intensities were converted to ranks across all 108 samples to account for uneven and non-normal distributions and due

to the fact that there is not a non-parametric equivalent of a two-way ANOVA. The two-way ANOVA was calculated on a per-metabolite basis. The resulting p-values were then corrected for multiple comparisons using False Discovery rate across all unidentified metabolites.

*Determination of Vitamin C Content (HPLC-PDA).* Extractions and analysis were performed according to published protocols[253] with modifications. Briefly, extraction solution comprising metaphosphoric acid (5% w/v) and dithiothreitol (DTT; 1% w/v) was added to potato samples and chilled at 4°C for 15 minutes. Chilled samples were homogenized at high speed in 5-10 second intervals to prevent heating and centrifuged for 15 minutes at 4°C and 1771 g. Supernatants were collected, filtered through 0.45 um nylon syringe filters into amber glass tubes, and stored at 4°C. Residues were re-extracted according to the above protocol without homogenization, and supernatants were stored in separate vials. Vitamin C analysis was performed on an HPLC (Shimadzu Prominence UFLC) with a refrigerated autosampler (4°C), column heater (30°C), and photodiode array detector (PDA). HPLC separation was achieved through a Zorbax SB-Aq column (Agilent, 3.5 um, 2.1 x 150 mm) using isocratic elution with 1% acetonitrile (ACN) / 20 mmol sodium phosphate buffer (pH 2). The flow rate was set at 0.2 mL/min, and the injected volume was 1 uL. The detector was set at 254 nm.

*Determination of Acrylamide Content (HPLC-PDA).* Extractions and analysis were

performed according to our established protocol (Ref) with minor modifications. Potato samples were extracted with methanol and centrifuged for 15 minutes at 10°C and 4000 g. Supernatants were treated with 50 uL each of Carrez I and Carrez II solutions and then centrifuged for 10 minutes at 10°C and 4000 g. Supernatant was transferred to glass test tubes and dried in a centrivap. Samples were immediately re-dissolved in 1 mL of ultra-pure water by vortex mixing for 2 minutes, filtered through 0.45 um nylon syringe filters, and stored at -20°C. Acrylamide analysis was performed on an HPLC (Agilent 1100) with a refrigerated autosampler (10°C), column heater (30°C), and photodiode array detector. HPLC separation was achieved through a Synergi 4u Hydro-RP 80A (Phenomenex, 5 um, 250 x 4.6 mm) using isocratic elution with 98% ultra-pure water (pH 6.5) and 2% acetonitrile. The flow rate was set at 2 mL/min, and the injected volume was 20 uL. The detector was set at 195 nm.

*Determination of Glycoalkaloid Content (HPLC-PDA).* Extractions and analysis were performed according to the established protocol (Ref) with minor modifications. Potato samples were extracted with acetone and centrifuged for 15 minutes at 4°C and 4000 g. Supernatants were transferred to chloroform resistant tubes. Residues were re-extracted with 5 mL aqueous acetone (30:70 acetone:water (v/v)) and centrifuged for 15 minutes at 4°C and 4000 g. Chloroform was added to the pooled supernatants (2:1 chloroform to extract), and stored overnight at 4°C. The top aqueous portion was filtered through 0.45 um nylon syringe filters, and

stored at -20°C. Glycoalkaloid (α –solanine and α-chaconine) analysis was performed on an HPLC (Agilent 1100) with a refrigerated autosampler (10°C), column heater (30°C), and photodiode array detector. HPLC separation was achieved through a Luna 5u C18(2) 100A column (Phenomenex, 5 um, 150 x 4.6 mm) using isocratic elution with acetonitrile and 0.05 M monobasic ammonium phosphate (35:65 v/v). The flow rate was set at 1 mL/ min, and the injected volume was 20 uL. The detector was set at 210 nm.

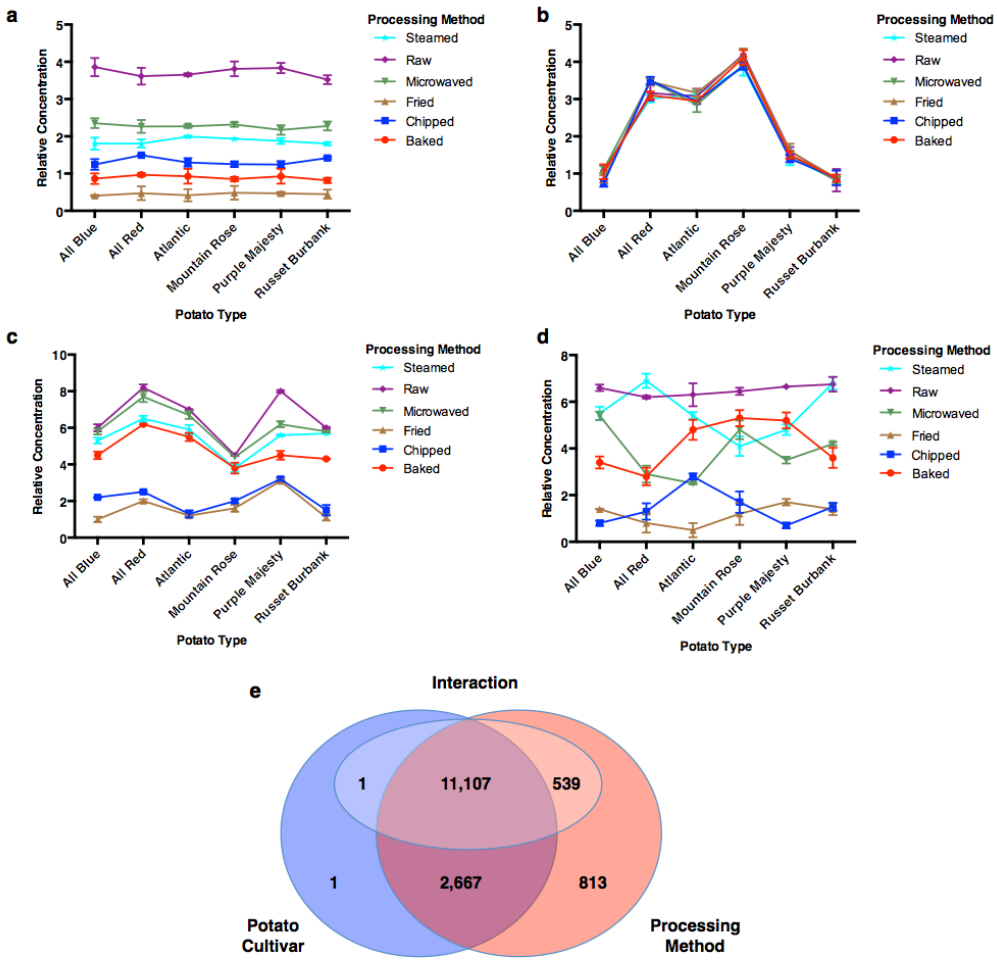


**Figure S6.1.** The effect of processing method, potato cultivar and interaction on the relative concentrations of potato metabolites. **a**, Example interaction plot where the relative concentration of a metabolite is only affected by processing method. **b**, Example interaction plot where the relative concentration of a metabolite is only affected by potato cultivar. **c**, Example interaction plot where the relative concentration of a metabolite is affected by potato cultivar and processing method, but with no interaction effect. In this case, the concentration of the metabolite abides by the additive model of the two variables. **d**, Example interaction plot where the relative concentration of a metabolite is affected by processing method, potato type, and an interaction effect. **e,** Each metabolite's relative abundance was converted to a rank abundance across all 108 samples to account for non-normal distribution. Two-way ANOVA was performed, testing for the effect of potato cultivar, processing method, and interaction. The Venn diagram contains the number of metabolites that were significantly different ($p < 0.05$ after False Discovery correction) due to the combination of these factors. Only 1 metabolite fragment was unaffected by potato cultivar and processing method. No metabolites were affected only by the interaction between potato cultivar and processing method. The majority of metabolites, 11,107 out of 15,129, were significantly

affected by potato cultivar, processing method, and the interaction between the two. Processing method affected more metabolites than potato cultivar, irrespective of interaction.

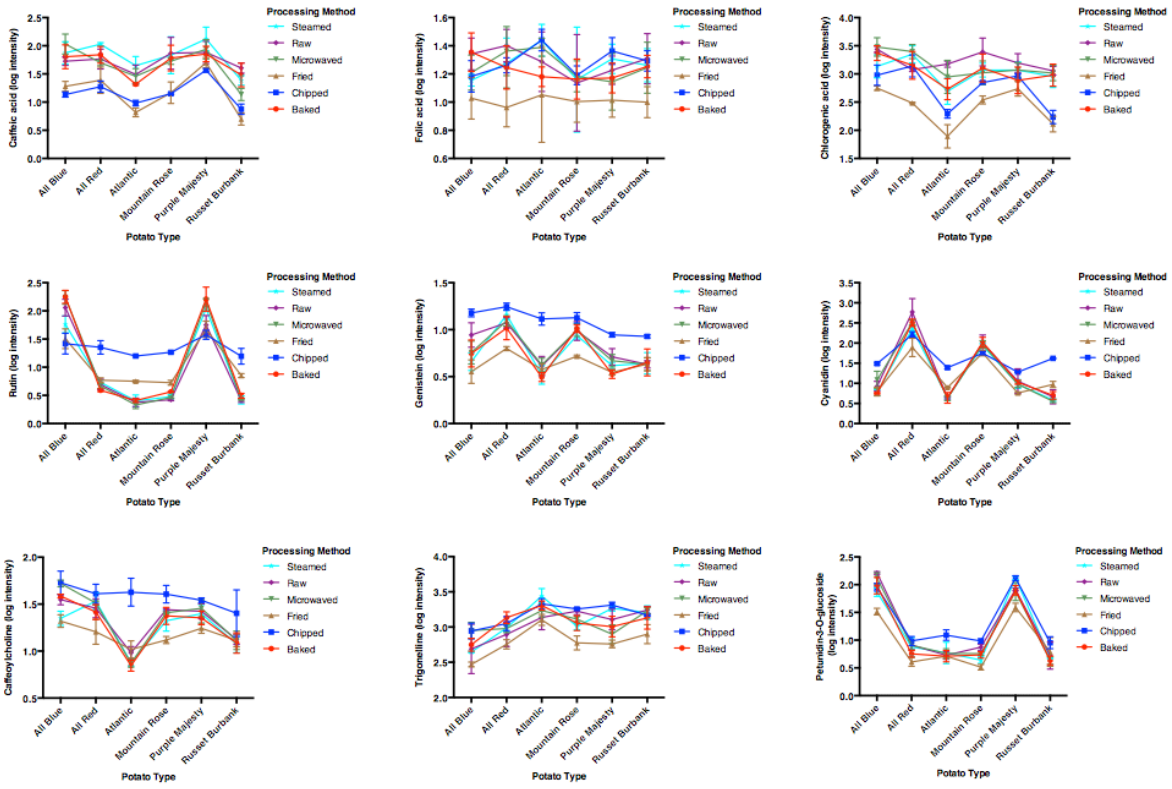


**Figure S6.2**. Interaction plots for targeted vitamins and antioxidants. The relative intensities for each identified metabolite were log transformed, and the average and standard deviation for replicates (n=3) of each treatment group are reported.

| | 15,129 Metabolites | | 9 Identified Metabolites | |
|---|---|---|---|---|
| Category | P-value | $R^2$ | P-value | $R^2$ |
| Potato Cultivar | 0.001 | 0.027 | 0.001 | 0.031 |
| Processing Method | 0.001 | 0.800 | 0.001 | 0.238 |
| Potato x Processing Interaction | 0.001 | 0.047 | 0.001 | 0.092 |

**Table S6.1. Results of two-way PERMANOVA testing the significance of the effects of potato cultivars, processing method and interaction effect, and explained variance.** The nine identified metabolites are caffeic acid, chlorogenic acid, folic acid, petunidin, rutin, genistein, cyanidin, caffeoylcholine, and trigonelline.

| | 15,129 Metabolites | | 9 Identified Metabolites | |
|---|---|---|---|---|
| Comparison | R | P-value | R | P-value |
| All | 0.073 | 0.002 | 0.694 | 0.001 |
| All Blue, All Red | 0.08 | 0.067 | 0.955 | 0.001 |
| All Blue, Atlantic | 0.084 | 0.067 | 0.888 | 0.001 |
| All Blue, Mountain Rose | 0.075 | 0.067 | 0.947 | 0.001 |
| All Blue, Purple Majesty | 0.016 | 0.257 | 0.157 | 0.004 |
| All Blue, Russet Burbank | 0.073 | 0.071 | 0.899 | 0.001 |
| All Red, Atlantic | 0.107 | 0.067 | 0.777 | 0.001 |
| All Red, Mountain Rose | 0.011 | 0.274 | 0.195 | 0.001 |
| All Red, Purple Majesty | 0.107 | 0.067 | 0.984 | 0.001 |
| All Red, Russet Burbank | 0.082 | 0.067 | 0.737 | 0.001 |
| Atlantic, Mountain Rose | 0.095 | 0.067 | 0.619 | 0.001 |
| Atlantic, Purple Majesty | 0.113 | 0.067 | 0.838 | 0.001 |
| Atlantic, Russet Burbank | 0.02 | 0.257 | 0.059 | 0.065 |
| Mountain Rose, Purple Majesty | 0.091 | 0.067 | 0.97 | 0.001 |
| Mountain Rose, Russet Burbank | 0.073 | 0.067 | 0.512 | 0.001 |
| Purple Majesty, Russet Burbank | 0.12 | 0.067 | 0.86 | 0.001 |

**Table S6.2. Results of an analysis of similarity (ANOSIM) test comparing differences in metabolite profiles between potato cultivars.** Bray-Curtis distances were calculated between all pairs of samples based on abundance of metabolites. Not accounting for processing methods, potato cultivars were compared using 999 permutations. The results of the global (All) and pairwise tests are shown. The 9 identified metabolites are caffeic acid, chlorogenic acid, folic acid, petunidin, rutin, genistein, cyanidin, caffeoylcholine, and trigonelline.

| | 15,129 Metabolites | | 9 Identified Metabolites | |
|---|---|---|---|---|
| Comparison | R | P-value | R | P-value |
| All | 0.884 | 0.001 | 0.214 | 0.001 |
| Baked, Chipped | 1 | 0.001 | 0.443 | 0.002 |
| Baked, Fried | 1 | 0.001 | 0.269 | 0.003 |
| Baked, Microwaved | 0.563 | 0.001 | -0.035 | 0.874 |
| Baked, Raw | 0.964 | 0.001 | -0.003 | 0.636 |
| Baked, Steamed | 0.197 | 0.004 | -0.029 | 0.874 |
| Chipped, Fried | 1 | 0.001 | 0.589 | 0.002 |
| Chipped, Microwaved | 1 | 0.001 | 0.399 | 0.002 |
| Chipped, Raw | 1 | 0.001 | 0.481 | 0.002 |
| Chipped, Steamed | 1 | 0.001 | 0.449 | 0.002 |
| Fried, Microwaved | 1 | 0.001 | 0.334 | 0.002 |
| Fried, Raw | 1 | 0.001 | 0.434 | 0.002 |
| Fried, Steamed | 1 | 0.001 | 0.343 | 0.002 |
| Microwaved, Raw | 0.876 | 0.001 | -0.039 | 0.874 |
| Microwaved, Steamed | 0.197 | 0.002 | -0.031 | 0.874 |
| Raw, Steamed | 0.944 | 0.001 | -0.037 | 0.874 |

**Table S6.3. Results of an analysis of similarity (ANOSIM) test comparing differences in metabolite profiles between processing methods.** Bray-Curtis distances were calculated between all pairs of samples based on abundances of metabolites. Not accounting for potato cultivar, processing methods were compared using 999 permutations. The results of the global (All) and pairwise tests are shown. The 9 identified metabolites are caffeic acid, chlorogenic acid, folic acid, petunidin, rutin, genistein, cyanidin, caffeoylcholine, and trigonelline.

| | 15,129 Metabolites | | 9 Identified Metabolites | |
|---|---|---|---|---|
| Comparison | R | P-value | R | P-value |
| All | 0.09 | 0.001 | 0.83 | 0.001 |
| Purple, Red | 0.085 | 0.005 | 0.957 | 0.001 |
| Purple, White | 0.094 | 0.005 | 0.867 | 0.001 |
| Red, White | 0.09 | 0.005 | 0.65 | 0.001 |

**Table S6.4. Results of an analysis of similarity (ANOSIM) test comparing differences in metabolite profiles between groups of potato flesh colors.** Bray-Curtis distances were calculated between all pairs of samples based on abundances of metabolites. Differences between color types were compared using 999 permutations. The results of the global (All) and pairwise tests are shown. The 9 identified metabolites are caffeic acid, chlorogenic acid, folic acid, petunidin, rutin, genistein, cyanidin, caffeoylcholine, and trigonelline.

| | 15,129 Metabolites | | 9 Identified Metabolites | |
|---|---|---|---|---|
| Comparison | R | P-value | R | P-value |
| All | 0.93 | 0.001 | 0.936 | 0.001 |
| All Blue, All Red | 0.852 | 0.001 | 1 | 0.001 |
| All Blue, Atlantic | 1 | 0.001 | 1 | 0.001 |
| All Blue, Mountain Rose | 0.994 | 0.001 | 1 | 0.001 |
| All Blue, Purple Majesty | 0.951 | 0.001 | 0.852 | 0.001 |
| All Blue, Russet Burbank | 0.914 | 0.001 | 1 | 0.001 |
| All Red, Atlantic | 0.932 | 0.001 | 1 | 0.001 |
| All Red, Mountain Rose | 0.772 | 0.001 | 0.679 | 0.001 |
| All Red, Purple Majesty | 0.944 | 0.001 | 1 | 0.001 |
| All Red, Russet Burbank | 0.87 | 0.001 | 1 | 0.001 |
| Atlantic, Mountain Rose | 1 | 0.001 | 1 | 0.001 |
| Atlantic, Purple Majesty | 1 | 0.001 | 1 | 0.001 |
| Atlantic, Russet Burbank | 0.895 | 0.001 | 0.481 | 0.001 |
| Mountain Rose, Purple Majesty | 1 | 0.001 | 1 | 0.001 |
| Mountain Rose, Russet Burbank | 0.926 | 0.001 | 1 | 0.001 |
| Purple Majesty, Russet Burbank | 0.926 | 0.001 | 1 | 0.001 |

**Table S6.5. Results of an analysis of similarity (ANOSIM) test comparing differences in metabolite profiles between potato cultivars within processing methods.** Bray Curtis distances were calculated between all pairs of samples based on abundances of metabolites. Accounting for processing methods, potato cultivars were compared using 999 permutations. The results of the global (All) and pairwise tests are shown. The 9 identified metabolites are caffeic acid, chlorogenic acid, folic acid, petunidin, rutin, genistein, cyanidin, caffeoylcholine, and trigonelline.

| | 15,129 Metabolites | | 9 Identified Metabolites | |
|---|---|---|---|---|
| Comparison | R | P-value | R | P-value |
| All | 0.972 | 0.001 | 0.741 | 0.001 |
| Baked, Chipped | 1 | 0.001 | 1 | 0.001 |
| Baked, Fried | 1 | 0.001 | 1 | 0.001 |
| Baked, Microwaved | 1 | 0.001 | 0.179 | 0.036 |
| Baked, Raw | 0.963 | 0.001 | 0.272 | 0.02 |
| Baked, Steamed | 0.926 | 0.001 | 0.364 | 0.003 |
| Chipped, Fried | 1 | 0.001 | 1 | 0.001 |
| Chipped, Microwaved | 1 | 0.001 | 1 | 0.001 |
| Chipped, Raw | 1 | 0.001 | 0.963 | 0.001 |
| Chipped, Steamed | 1 | 0.001 | 1 | 0.001 |
| Fried, Microwaved | 1 | 0.001 | 1 | 0.001 |
| Fried, Raw | 1 | 0.001 | 0.975 | 0.001 |
| Fried, Steamed | 1 | 0.001 | 1 | 0.001 |
| Microwaved, Raw | 0.92 | 0.001 | 0.34 | 0.001 |
| Microwaved, Steamed | 0.994 | 0.001 | 0.401 | 0.006 |
| Raw, Steamed | 0.963 | 0.001 | 0.463 | 0.001 |

**Table S6.6. Results of an analysis of similarity (ANOSIM) test comparing differences in metabolite profiles between processing methods within potato cultivars.** Bray Curtis distances were calculated between all pairs of samples based on abundances of metabolites. Accounting for potato cultivar, processing methods were compared using 999 permutations. The results of the global (All) and pairwise tests are shown. The 9 identified metabolites are caffeic acid, chlorogenic acid, folic acid, petunidin, rutin, genistein, cyanidin, caffeoylcholine, and trigonelline.