HIV-Associated Gut Microbiome Differences and Immune Activation are Dependent on Host Context

HIV-associated Gut Microbiome Differences and Immune Activation are Dependent on Host Context

A dissertation presented

by

David B. Gootenberg

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Virology

Harvard University

Cambridge, Massachusetts

April 2019

Dissertation Advisor: Dr. Douglas S. Kwon David B. Gootenberg

HIV-associated Gut Microbiome Differences and Immune Activation are Dependent on Host Context

Abstract

Even with antiretroviral therapy, individuals infected with Human Immunodeficiency Virus (HIV) experience greater mortality than uninfected individuals. This mortality is due in large part to inflammation-related non-communicable diseases, such as cardiovascular disease and kidney dysfunction, but the causes of the underlying inflammation are not fully understood. One proposed driver is the gut microbiome, which is essential to human health and involved in many metabolic and immune interactions. Recent studies have found HIV-associated disruptions in the gut microbial community, but few of these data are derived from subjects in sub-Saharan Africa where HIV burden is greatest. We hypothesized that HIV-associated gut microbiome differences vary by geography, influencing the relationship with elevated host inflammation. To test this, we profiled the gut microbiota and measured serum inflammatory markers in more than 700 subjects in rural Mbarara, Uganda; urban Gaborone, Botswana; and Boston, USA. Using computational and statistical methods, we found that within each geographic cohort, the microbiomes of HIV-infected and -uninfected individuals exhibited significant differences. In the Ugandan cohort, HIV-uninfected individuals had communities dominated by the genus *Prevotella*, while HIV-infected individuals had relatively lower abundance of *Prevotella* and a greater abundance of taxa including *Bifidobacterium* and *Akkermansia*. In contrast, gut

communities from HIV-infected individuals in Boston showed increased *Prevotella* and decreased *Akkermansia* and other taxa relative to HIV-uninfected subjects. Compared to the other cohorts, individuals in the Botswanan cohort had both common and distinct HIV-associated differences. Within the Boston cohort, sexual practices were an interacting factor as men who have sex with men (MSM) exhibited greater HIV-associated gut microbiome differences. Inflammatory markers were elevated in HIV-infected individuals in all cohorts and this inflammation was associated with unique bacterial taxa in each cohort. We isolated bacteria within the species *Prevotella copri* from individuals across the cohorts and found geographically associated functional differences, demonstrating that host context can contribute to microbial differences both at high taxonomic levels and within sub-species diversity. In order to most effectively translate HIV-gut microbiome research into opportunities for therapeutic interventions, studies should include populations in multiple geographic locations, as we found HIV-associated changes to be highly context dependent.

Table of Contents

Attribution of Collaborator Contributions

Chapter 1: David B. Gootenberg (DBG) wrote this section incorporating text and figures adapted from (Gootenberg et al., 2017) and (Anahtar et al., 2018) that were originally produced by DBG.

Chapter 2: DBG performed the experimental work with the assistance and guidance of Cynthia L. Monaco (CLM). CLM planned the experiments and analyzed the data with the assistance of Musie S. Ghebremichael and DBG. Mark J. Siedner (MJS) established the clinical cohort and assisted in clinical data curation. Jesús Mario Luévano (JML) and Meaghan T. Flagg assisted DBG with fecal nucleic acid extractions and 16S rRNA gene amplicon generation. CLM and DBG collaborated in writing the original published manuscript. DBG adapted the manuscript into this section of the dissertation with input and guidance from Douglas S. Kwon (DSK).

Chapter 3: DBG planned and performed the majority of the experimental work in this section. Assistance was given for the following portions of the work: fecal nucleic acid extractions and 16S rRNA gene amplicon generation: Jeffrey M. Paer, JML, and Zoe H. Rogers; soluble inflammatory marker quantification: University of Vermont Department of Pathology; *Prevotella* isolation: Seth M. Bloom and Juliet T. Bramante (JTB); *Prevotella* sequencing: Jiawu Xu (JX); assembly of *Prevotella* genomes: Matthew R. Hayward (MRH). DBG wrote this section with input and guidance from DSK.

Chapter 4: DBG wrote this section with input and guidance from DSK.

Chapter 5: DBG wrote this section with input and guidance from DSK. Additional methods language was provided by CLM, JML, JTB, JX, and MRH.

Acknowledgements

The work covered in this dissertation and all that I accomplished during graduate school would not have been possible without the assistance, mentorship, and support of far too many people to cover in these acknowledgements. To do justice to the contributions made by this community to this work and my scientific maturation would require a second dissertation all on its own. Know that I appreciate every bit of help I received along this journey, no matter how large or small.

Scientific training is an endeavor with many parallels to apprenticeship. As such, I was extremely fortunate to conduct my training under the tutelage of Doug Kwon. Doug was integral to the genesis of this work and my graduate career. I cannot thank him enough for putting his faith in a medical student who decided midway through his training to add a graduate degree. Doug shepherded me through countless grant applications, professional interactions, and public scientific presentations that I would not have been able to tackle on my own. He guided me through my scientific adolescence with a steady hand and patience that resulted in the scientific adulthood culminating in the thesis presented here. It was a privilege to have you as my advisor and I know that for the rest of my career I will be part of your scientific lineage.

Not only was I fortunate to work with Doug, but I was able to find a home within the Kwon laboratory as a whole. In my nearly 5 years in the laboratory, many people passed through the lab, each contributing in their own way to my experience in graduate school. I must first thank those graduate students who shared nearly the entire journey with me: Crystal Rawlings, Abigail Schiff, and Meaghan Flagg. I could not have hoped to have had a better group of compatriots to endure graduate school with. Your

kindness, friendship, support, and insight were invaluable over the half-decade we spent together and I hope that I was able to repay in kind even of a fraction of what you gave me. I cannot give enough thanks to the members of the “Poop Group” who helped me complete the work on the gut microbiome that comprises this thesis: Jesús Mario Luévano, Jeff Paer, Zoe Rogers, Alex Muñoz, and our newcomer Muntsa Rocafort. For Jeff and Jesús, the opportunity to mentor you was one I will remember fondly for the rest of my life. Muntsa, I cannot wait to see the next steps in this work. I also have to thank those “Honorary Poop Group” members who pitched in to complete a work that covered an enormous scale: Matt Hayward, Jiawu Xu, Seth Bloom, and Juliet Bramante. Whether your assistance was *in silico*, *in vivo*, or both, it would have been impossible to finish this project without you. On the clinical dimension of this work, I want to thank the clinical coordinators that worked tirelessly to recruit subjects for this study: Daniel Worrall, Sheila Ip, Frank Ruzicka, Lyan Chung, and Olivia Carr. Lastly I want to thank those clinical subjects who participated in our studies across at least three countries and two continents without which this work literally could not have been done.

I want to thank all those in the wider scientific world who leant me a hand during my time in graduate school. Our collaborators, Mark Siedner and Mosepele Mosepele, for allowing us to work with you on clinical cohorts that you painstakingly planned, recruited, and maintained. Members of the Virgin Lab at Washington University in St. Louis who provided me with so much scientific assistance: Cynthia Monaco, Scott Handley, Guoyan Zhao, Megan Baldridge, and “Skip” Virgin himself. The Ragon Bioinformatics Working Group at the Ragon Institute that contained some of my favorite partners in crime on quantitative and computational work, especially Jonathan Urbach

and Musie Ghebremichael. I also want to acknowledge the generous funding from the National Heart, Lung, and Blood Institute and the HIV Medical Association that supported my work.

I want to thank all the institutions and programs that provided the infrastructure to support me as a graduate student. The MD-PhD and Virology programs were an irreplaceable source of assistance and wisdom. I could not imagine a better program than HST in which to receive my medical training and they have helped me in all my accomplishments and struggles along the way. Within HST, I would like to specifically acknowledge Patty Cunningham, Matthew Frosch, and Rick Mitchell for their help and support. I want to acknowledge my larger scientific home of the Ragon Institute and its director, Bruce Walker, who nurtured an environment full of scientific colleagues who contributed so much to my scientific growth. I received invaluable scientific assistance and mentorship from my preliminary qualifying exam committee: Max Nibert, Daniel Kuritzkes, and Pardis Sabeti; my dissertation advisory committee: Todd Allen, Eric Alm, Seth Rakoff-Nahoum, and Pardis Sabeti; and my dissertation defense committee: Todd Allen, Kate Jeffrey, Alex Kostic, and Tami Lieberman.

I must thank the many friends and classmates who travelled this academic journey with me. My high school friends, who were with me at the first spark of my interest in science and medicine and have been with me ever since, including one of my oldest and closest friends Pawel Przytycki, who along with George Moomau and Mantas Vidutis schlepped up to Boston on a Wednesday with 6 weeks notice to attend my wedding. My friends from college, especially those in the Harvard Band, many of them who are still in Boston and continued to interact with me socially even as I

descended into the deeper reaches of “Graduate-student-hood.” My medical school friends with whom I began the first steps of my career and then watched grow from students into physicians. My graduate school friends from the Virology program who comprised a tight-knit group of fellow travelers that I am glad to have shared some of the most trying years of my graduate school journey with. I also must acknowledge my colleagues at the Harvard Graduate Student Union who worked tirelessly on behalf of graduate students long before I entered graduate school and long after I have graduated.

I want to thank my family for their support during my time in graduate school. I am so fortunate to have Jonathan Gootenberg as a younger brother who is both my scientific confidant and scientific inspiration. Your success is both unbelievable and well deserved. My parents, Joseph Gootenberg and Susan Leibenhaut, have been the most steady, loving, caring, wise, and inspiring source of support my entire life. I cannot put it better than the acknowledgments of my brother’s dissertation, which read: “I would like to thank them for guiding me to be the best person I can be, imbuing me from a young age with a curiosity and drive for science, and for teaching me that helping the world is of utmost importance.” I also deeply appreciate their understanding attitude towards the long periods in which I was incommunicado due to work on this project.

Lastly, I want to thank my wife, Larissa Wenren, who has touched every facet of my life and in doing so elevated each of them beyond description. She has stood by me with unwavering support during the difficult journey of graduate school and in a way has earned her own place as a co-author of this work. To even attempt to thank her appropriately, I must borrow from my wedding vows:

“Larissa Wenren, you absolutely astound me with your skill and accomplishments. You have a heart so bright that it spreads warmth to everyone you meet. Your patience is seemingly infinite, and the laughs we have shared have brought me my greatest moments of joy. There is no obstacle that you have not been able to outthink, outwit, or outwork. Every time I think you have risen to an unsurpassable level as a partner and a person, you keep reaching higher. You are my best partner and my best friend. The person who makes me feel the most safe and loved. I want to spend the rest of my life with you. I want to build a home and a family and have the chance to share every single meaningful moment with you.”

Larissa: this dissertation, like so much else in my life, would not have been possible without you.

Chapter 1: Introduction

1.1 The human enteric microbiota

The human gastrointestinal tract harbors a diverse ecosystem of trillions of Bacteria, Archaea, Eukarya, and viruses (Delwart, 2013; Ley et al., 2006b; Norman et al., 2014; Virgin, 2014; Virgin and Todd, 2011; Virgin et al., 2009; Wylie et al., 2012), which is collectively called the enteric "microbiota" and contains a large aggregate genome, referred to as the "microbiome." Prior work in the gut, which is one of the most studied human-associated microbial environments (Human Microbiome Project, 2012), has shown that the enteric microbiota are essential participants in many host processes, both physiological (Honda and Littman, 2012) and pathological (Cho and Blaser, 2012; Friedrich, 2013; Littman and Pamer, 2011). The gut microbiome contains numerous genes that enable the synthesis of essential amino acids, methane, vitamins, and isoprenoids, in addition to the metabolism of many otherwise indigestible components of our diet (Gill et al., 2006; Sousa et al., 2008). The gut microbiota is an important factor in host energy balance (the equilibrium between caloric intake and expenditure), as different gut microbiome compositions have been found to be responsible for differences in metabolic efficiency and ability to extract energy from food, and data indicate that these variations can influence whether an individual is obese or lean (Peterson and Turnbaugh, 2010; Turnbaugh et al., 2008; Turnbaugh et al., 2006; Vijay-Kumar et al., 2010; Wen et al., 2008). Conversely, host diet itself can substantially affect the composition and function of the microbiota (David et al., 2014; Minot et al., 2011; Wu et al., 2011).

With regards to pathology, the gut microbiome has been associated with a variety of human diseases (Cho and Blaser, 2012; Friedrich, 2013; Littman and Pamer, 2011), including metabolic syndrome (Turnbaugh et al., 2006; Vijay-Kumar et al., 2010;

Wen et al., 2008), inflammatory bowel disease (IBD) (Cadwell et al., 2010; Gevers et al., 2014; Kostic et al., 2014; Mazmanian et al., 2008; Peterson and Cardona, 2010; Peterson and Turnbaugh, 2010; Scher et al., 2015; Willing et al., 2010), and even gastrointestinal cancers (Kostic et al., 2013; Vannucci et al., 2008). In general, high alpha diversity communities in the gut have been associated with a healthy host while low alpha diversity has been described in a number of diseases (Backhed et al., 2012; Lyte, 2013; Norman et al., 2015), such as IBD (Gevers et al., 2014; Kostic et al., 2014; Scher et al., 2015; Willing et al., 2010), type-1 diabetes mellitus (Kostic et al., 2015), obesity (Le Chatelier et al., 2013; Sze and Schloss, 2016), and *Clostridium difficile* infection (Chang et al., 2008; Weingarden et al., 2015), though many other pathologies have different relationships with gut community diversity.

Because the human microbiome has the potential to affect so many aspects of human health, it has been the focus of a series of international human microbiome projects (Nelson et al., 2010; Qin et al., 2010; Turnbaugh et al., 2007). Laboratories engaged in these human microbiome projects are harnessing the falling costs and increasing throughput of nucleic acid sequencing to characterize the organisms, genes, and functional abilities found in microbial communities across many diverse body habitats (Turnbaugh et al., 2007). These endeavors are analogous to the preceding Human Genome Project in that they seek to provide a comprehensive overview of one aspect of human biology that significantly influences human health.

1.2 Relationship between the enteric microbiota and host immune system

The host immune system acts as an essential curator for this luminal enteric microbial community, serving to shape and control the structure and function of this

diverse collection of organisms (Hooper et al., 2012; Thaiss et al., 2016; Zhang and Luo, 2015). Antimicrobial peptides like the REG3 family of C-type lectins and host secreted immunoglobulin A (IgA) preferentially select for certain bacteria within gut microbial communities (Peterson et al., 2007; Vaishnava et al., 2011). Small molecules are also part of the host arsenal, as secreted bile acids act as bactericidal surfactants and are able to selectively inhibit susceptible organisms and change enteric microbial composition (Islam et al., 2011). The growth of desired commensal organisms can also be supported through production of a wide variety of carbohydrates present in host glycans (Koropatkin et al., 2012), and host secretion of specific carbohydrates in forms such as fucosylated proteins (Pickard et al., 2014).

While the host immune system influences the gut microbial community, there is also a reciprocal process of resident microbes modulating the host immune milieu that contributes to physiological immune development (Honda and Littman, 2012). In one example of bacterial action on the host, members of the genera *Bacteroides, Bifidobacterium, and Lactobacillus* as well as the class Clostridia have been shown to induce the differentiation of regulatory T cells (T_{regs}) in the gut mucosa (Atarashi et al., 2013; Atarashi et al., 2011; Furusawa et al., 2013; Tanoue et al., 2016). These cells produce anti-inflammatory cytokines such as interleukin-10 (IL-10) and dampen host inflammation (Donaldson et al., 2016; Tanoue et al., 2016). Some of these bacterial-associated effects are mediated by secreted bacterial products, including the capsule component polysaccharide A (PSA) from *Bacteroides fragilis* (Round and Mazmanian, 2010) and short-chain fatty acid (SCFA) metabolites produced by members of the class Clostridia (Atarashi et al., 2013).

The gut microbiota can also dramatically impact the development of other T helper lineages. In mice, segmented filamentous bacteria (SFB) of the class Lachnospiraceae have been shown to regulate expansion of the enteric Th17 T cell compartment, which is important for homeostasis of the immune-microbiota interface but can also lead to autoimmune disease in certain contexts (Ivanov et al., 2009). These examples of host immune modulations mediated by gut microbiota, often through secreted bacterial products, provide motivation for searching for similar immunomodulatory mechanisms, as these mechanisms could be leveraged to improve human health.

1.3 Chronic systemic immune activation in HIV Infection

Globally, nearly 37 million people are infected with human immunodeficiency virus-1 (HIV), and this population is highly concentrated in sub-Saharan Africa (UNAIDS, 2018). HIV targets and destroys host CD4+ T cells, eventually progressing into the development of acquired immunodeficiency syndrome (AIDS), which is reached when a patient has fewer than 200 CD4+ T cells/μl present in their systemic circulation or they develop an AIDS-defining opportunistic infection or cancer (Selik et al., 2014).

Despite advancements in HIV antiretroviral therapy (ART), average life expectancy of HIV-infected individuals on treatment is significantly less than that of uninfected persons (**Table 1.1**) (Collaboration of Observational et al., 2012; Egger et al., 2002; Katz and Maughan-Brown, 2017). The ART Cohort Collaboration Study found that deaths were due largely to inflammation-related non-communicable diseases (NCD), such as stroke, long-bone fractures, cardiovascular disease (CVD), and renal dysfunction (Centers for Disease Control and Prevention, 2008; Egger et al., 2002;

Lewden et al., 2007; Lohse et al., 2007; Triant et al., 2008). In a cohort of HIV-infected Ugandans, a 1.6 fold increased hazard of death was associated with each 10% increase in CD8+ T-cell activation following initial viral suppression with ART (Hunt et al., 2011). HIV infection also frequently correlates with increased tissue and circulating measures of inflammation (*e.g.* circulating soluble CD14 [sCD14] and IL-6, tissue CD38+HLA-DR+CD8+ T-cells) (Brenchley et al., 2006; Cassol et al., 2010; Sandler et al., 2011) and these measures of inflammation are associated with mortality in multiple populations (**Tables 1.2** and **1.3**) (Hunt et al., 2014; Lee et al., 2017). CVD, which constitutes a large proportion of HIV-associated NCD (Muyanja et al., 2015; Okello et al., 2015; Siedner et al., 2016a), is increased in HIV infection (**Table 1.4**) (Chow et al., 2012; Freiberg et al., 2013; Grunfeld et al., 2009; Marcus et al., 2014; Triant et al., 2007) and associated with systemic immune activation as measured by markers such as serum IL-6, soluble CD163 (sCD163) and C-reactive protein (CRP) (Duprez et al., 2012; McKibben et al., 2015; Ross et al., 2009; Subramanian et al., 2012; Triant and Grinspoon, 2011; Triant et al., 2009).

Further supporting the connection between systemic inflammation and chronic HIV pathogenesis, individuals with the highest degree of persistent elevated immune activation while on suppressive ART experience higher overall mortality, even with CD4+ T-cell reconstitution greater than 500 cells/μl (Hunt, 2012; Tien et al., 2010). Although systemic immune activation declines after initiation of ART, it remains persistently elevated compared to HIV-uninfected subjects in the majority of subjects even after years of therapy (Gandhi et al., 2006; Goicoechea et al., 2006; Hunt et al.,

2003) and has also been observed in individuals with undetectable viral loads (Hunt et al., 2008).

Table 1.1: Increased mortality rate in ART-treated HIV-infected individuals. Standardized mortality ratio (observed/expected deaths) and excess mortality rates in HIV-infected individuals with CD4+ T cell count greater than 500/µl after ART initiation according to gender and age. SMR was significantly greater than one for all but the oldest age groups. PY: person-years, SMR: standardized mortality ratio, n = 35,516. Adapted from (Collaboration of Observational et al., 2012).

	PY	Mortality rate/ 100 PY (95% CI)	SMR (95% CI)	Excess mortality rate/ 100 PY (95% CI)
Global	91 891	0.37 (0.30–0.46)	1.5 (1.2–1.8)	0.12 (0.10–0.15)
Men (years)	65 697	0.43 (0.35–0.53)	1.4 (1.1–1.7)	0.12 (0.10-0.15)
18–39	32 565	0.31 (0.24–0.40)	2.5 (1.9–3.2)	0.18 (0.14-0.24)
40–59	30 216	0.48 (0.38–0.62)	1.3 (1.0–1.7)	0.12 (0.09–0.15)
⩾60	2916	1.27 (0.88–1.82)	0.7 (0.5–1.0)	
Women (years)	26 194	0.23 (0.17–0.32)	2.2 (1.6–3.0)	0.13 (0.09–0.17)
18–39	17 564	0.17 (0.12–0.24)	3.4 (2.5–4.8)	0.12 (0.09–0.17)
40–59	7938	0.32 (0.23–0.45)	1.9 (1.3–2.6)	0.15 (0.10–0.21)
⩾60	692	0.77 (0.50–1.21)	1.0 (0.6–1.5)	−0.03 (−0.04 to 0.02)

Table 1.2: Soluble biomarkers of inflammation predict mortality in American Cohort. Soluble biomarker predictors of mortality among 192 participants in the Longitudinal Study of the Ocular Complications of AIDS who had ART-suppressed HIV infection. All four biomarkers shown are associated with a significant increase in odds-ratio (OR) of death. Adapted from (Hunt et al., 2014)

	OR (95% CI) for Death, by Quartile[b]						OR per IQR	
Characteristic, Analysis[a]	Second	*P* Value	Third	*P* Value	Fourth	*P* Value	Increase (95% CI)	*P* Value
Proximal CD4$^+$ T-cell count, cells/mm^3								
Primary	0.50 (.22–1.1)	.099	0.41 (.17–.98)	.045	0.44 (.18–1.1)	.076	0.62 (.40–.95)	.030
I-FABP level, pg/mL								
Primary	1.76 (.61–5.1)	.30	4.5 (1.5–13.3)	.007	8.3 (2.8–25.1)	<.001	3.5 (2.0–6.1)	<.001
Adjusted	1.69 (.56–5.1)	.35	4.2 (1.4–12.8)	.011	8.6 (2.7–27.8)	<.001	3.5 (1.9–6.1)	<.001
sCD14 level, µg/mL								
Primary	2.7 (.82–9.1)	.10	7.7 (2.3–25.7)	.001	17.6 (4.4–55.1)	<.001	5.4 (2.8–10.4)	<.001
Adjusted	4.5 (1.09–18.6)	.038	11.4 (2.9–46)	.001	30.1 (6.2–145)	<.001	7.5 (3.4–16.5)	<.001
KT ratio, nM/µM								
Primary	1.48 (.51–4.3)	.47	2.3 (.82–6.5)	.11	4.6 (1.72–12.3)	.002	2.3 (1.45–3.5)	<.001
Adjusted	1.50 (.50–4.4)	.47	2.4 (.82–6.9)	.11	4.3 (1.51–12.4)	.006	2.3 (1.40–3.7)	.001
IL-6 level, pg/mL								
Primary	6.4 (1.33–30.6)	.020	9.8 (1.89–50.5)	.007	69.7 (12.4–392)	<.001	6.1 (2.9–12.9)	<.001
Adjusted	12.0 (1.42–102)	.023	17.8 (2.1–154)	.009	139 (14–1362)	<.001	6.6 (2.9–15.0)	<.001

Table 1.3: Soluble biomarkers of inflammation predict mortality in Ugandan Cohort. Predictors of late mortality among 457 HIV-infected Ugandans with a plasma HIV RNA load less than 400 copies/mL by month 6 of ART. Multiple biomarkers, including IL-6 and sCD14, were associated with a significantly increased hazard ratio (HR) of death using different regression models. From (Lee et al., 2017).

Biomarker (Units)	Univariate HR (95% CI)[a]	Multivariate aHR I (95% CI)[b]	Multivariate aHR II (95% CI)[c]
KT ratio[d,e]	2.43 (1.62–3.65)	2.74 (1.63–4.60)	…
D-dimer (μg/mL)	1.87 (1.00–3.53)	1.95 (1.00–3.80)	1.83 (.82–4.11)
IL-6 (pg/mL)	2.16 (1.51–3.08)	2.34 (1.51–3.63)	1.84 (1.10–3.05)
sCD163 (μg/L)	1.92 (.98–3.76)	1.52 (.73–3.19)	0.97 (.43–2.22)
sCD14 (μg/L)	2.19 (1.28–3.75)	2.32 (1.25–4.31)	1.54 (.76–3.12)
I-FABP (pg/mL)	1.01 (.51–2.02)	0.89 (.44–1.79)	0.73 (.36–1.50)
$CD4^+$ T-cell activation[f]	1.95 (1.07–3.56)	2.04 (.95–4.42)	1.64 (.75–3.57)
$CD8^+$ T-cell activation[g]	1.96 (.99–3.87)	2.09 (.93–4.68)	1.79 (.80–3.99)
$CD4^+/CD8^+$ ratio[h]	0.63 (.39–1.04)	1.08 (.55–2.12)	1.03 (.46–2.27)

Table 1.4: HIV independently increases CVD risk. HIV infection and the risk of acute myocardial infarction (MI). HIV infection was independently associated with a hazard ratio of acute MI on the scale similar to that measured from hypertension, diabetes mellitus, and smoking. Total n = 82,459 (55,109 HIV-uninfected, 27,350 HIV-infected). Adapted from (Freiberg et al., 2013).

Characteristic	Acute MI Hazard Ratio (95% CI)
HIV infection	1.48 (1.27-1.72)
Hypertension	
None	1 [Reference]
Controlled	1.36 (1.08-1.70)
Uncontrolled	1.64 (1.41-1.91)
Diabetes mellitus	1.74 (1.49-2.02)
Never smoker	1 [Reference]
Current smoking	1.78 (1.47-2.16)
Past smoking	1.06 (0.80-1.40)

1.4 Effects of HIV infection on the enteric microbiome

HIV infection leads to the widespread destruction of host immune function (Andrews and Koup, 1996; Okoye and Picker, 2013), including the rapid and profound depletion of CD4+ T-cells within gut associated lymphoid tissue (GALT) (Brenchley et al., 2004; Douek et al., 2009; Klatt et al., 2008; Sandler and Douek, 2012). Early in HIV infection the GALT is a critical reservoir of viral replication and substantial CD4+ T-cell depletion (80-90%) (Brenchley et al., 2004; Clayton et al., 1997; Heise et al., 1994; Li et al., 2005; Mattapallil et al., 2005; Mehandru et al., 2004; Schneider et al., 1996; Schuetz et al., 2014; Veazey et al., 1998). This depletion is attributed to the widespread surface expression of CCR5 on gastrointestinal T-cells, which serves as the coreceptor for HIV entry early on in infection (Brenchley et al., 2004; Okoye and Picker, 2013). Even after successful control of systemic HIV replication with ART, immune recovery in the GI tract can be variable (Maartens et al., 2014).

As could be predicted from the loss of mucosal immune cells that normally serve to maintain homeostasis with the microbiota (Hooper et al., 2012; Hooper and Macpherson, 2010; Pickard et al., 2014), changes in the enteric microbial community have been reported during HIV infection (**Table 1.5**). Comparative studies between HIV-infected and uninfected individuals have differed in methods and patient populations studied (**Table 1.6**) and found a wide spectrum of differences associated with HIV infection (**Table 1.5**). However, several overarching themes have emerged. HIV-infected individuals often have reduced alpha diversity of their enteric microbiome, which, as previously mentioned, is also observed in a number of pathologies such as IBD (Gevers et al., 2014; Kostic et al., 2014; Scher et al., 2015; Willing et al., 2010), type-1 diabetes mellitus (Kostic et al., 2015), obesity (Le Chatelier et al., 2013; Sze and

Schloss, 2016), and *Clostridium difficile* infection (Chang et al., 2008; Weingarden et al., 2015). Also frequently observed are increases in traditionally pathogenic bacteria such as Enterobacteriaceae (Bender et al., 2016; Dinh et al., 2015; Dubourg et al., 2016; Ling et al., 2016; Monaco et al., 2016; Mutlu et al., 2014; Vujkovic-Cvijin et al., 2013) and decreases in commensals such as Lactobacillaceae, Lachnospiraceae, and Ruminococcaceae (Dillon et al., 2014; Dillon et al., 2016; Dubourg et al., 2016; Ellis et al., 2011; Gori et al., 2008; McHardy et al., 2013; Monaco et al., 2016; Mutlu et al., 2014; Nowak et al., 2015; Sun et al., 2016; Vázquez-Castellanos et al., 2015; Vujkovic-Cvijin et al., 2013; Yu et al., 2014). Many studies have found increases in members of the genus Prevotella in HIV-infected individuals and speculate that this is associated with an elevated inflammatory state (Lozupone et al., 2013; Lozupone et al., 2014). Further detail about studies conducted on the gut microbiome in HIV is provided in **Appendix 1**.

Table 1.5. Alterations in the gut microbiota reported in HIV microbiome studies
Bracketed letter indicates taxonomic level of classification (e.g. [p] indicates phylum), parenthetical taxa indicate subtaxa that commonly drive significance of parent taxa.
[a]Proteobacteria class indicated by corresponding Greek letter (e.g. β for Betaproteobacteria).
[b]Dashes denote studies that did not report alpha diversity.
[c]Microbial changes observed in HIV positive subjects with less than 200 CD4+ T cells/μL.
Adapted from (Gootenberg et al., 2017)
Data assembled from (Armstrong et al., 2018; Dillon et al., 2017; Dillon et al., 2014; Dillon et al., 2016; Dinh et al., 2015; Dubourg et al., 2016; Hoel et al., 2018; Lee et al., 2018; Ling et al., 2016; Liu et al., 2019; Lozupone et al., 2013; Lozupone et al., 2014; Lu et al., 2018; Maurice et al., 2019; McHardy et al., 2013; Monaco et al., 2016; Mutlu et al., 2014; Neff et al., 2018; Noguera-Julian et al., 2016; Nowak et al., 2015; Nowak et al., 2017; Pinto-Cardoso et al., 2017; Qing et al., 2018; San-Juan-Vergara et al., 2018; Serrano-Villar et al., 2017; Sessa et al., 2019; Sun et al., 2016; Vazquez-Castellanos et al., 2018; Vázquez-Castellanos et al., 2015; Vesterbacka et al., 2017; Villanueva-Millan et al., 2017; Vujkovic-Cvijin et al., 2013; Williams et al., 2019; Yang et al., 2016; Yu et al., 2014; Zhou et al., 2018b)

Table 1.5 (Continued)

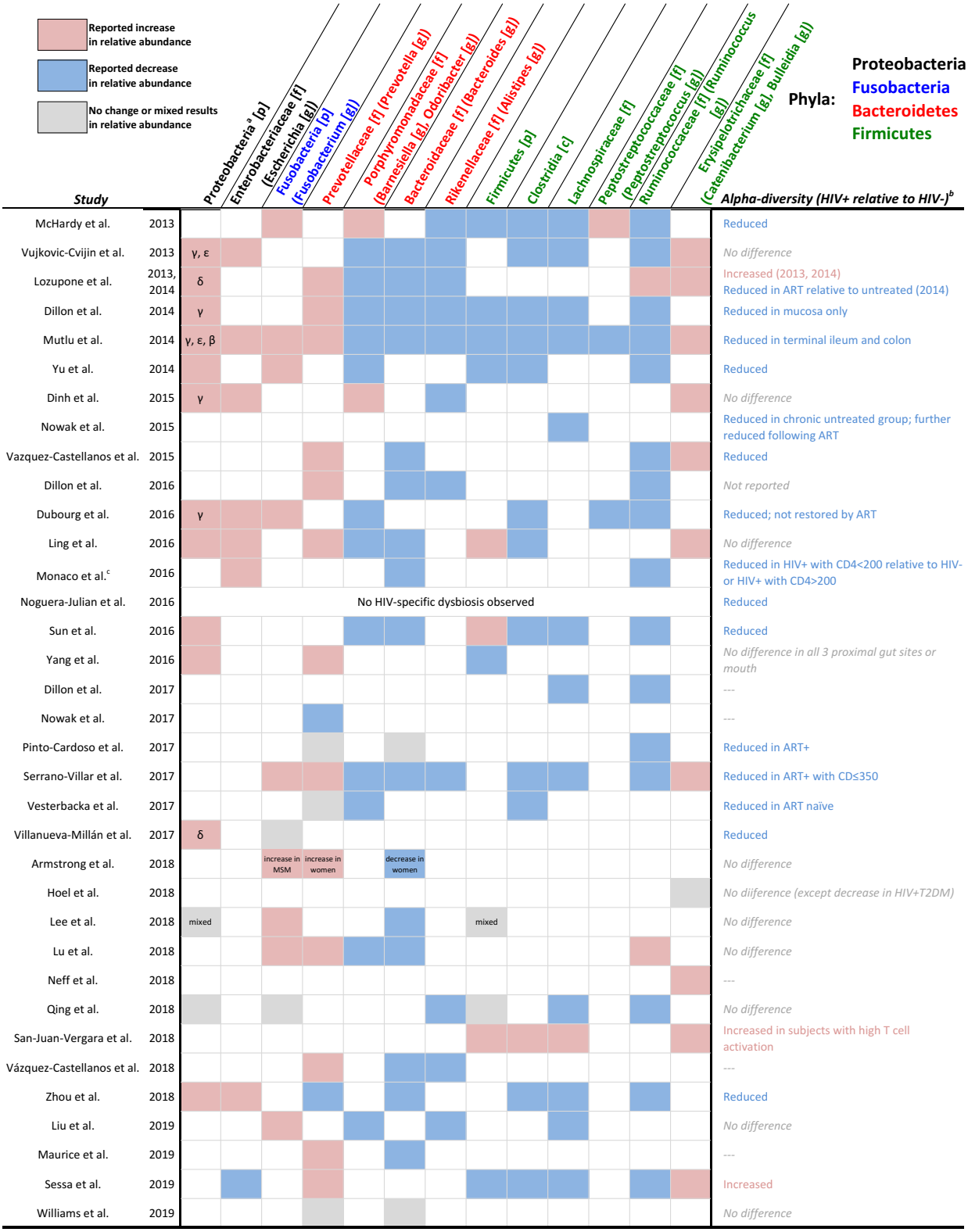

Table 1.6. Study design and patient metadata in HIV microbiome studies.
*CD4 counts (cells/μL) displayed as median (range [not notated in table]) median (IQR), or mean±standard deviation. [†]Sex composition of clinical subject groups not specified.
[a]The same patient cohort was used for both (Lozupone et al., 2013; Lozupone et al., 2014). Displayed is patient information from (Lozupone et al., 2014), which included 14 additional subjects on ART, 1 chronic progressor, and 2 HIV negative controls. Proportion of males versus females in each clinical group is only described in (Lozupone et al., 2013) (male/female: 8/5 seronegative; 11/0 HIV+ untreated; 7/1 HIV+ treated).
[b]Fecal microbiome data only obtained for HIV seronegative unexposed and exposed infants. Bender et al. 2016 included 100 total subjects (50 mother-infant pairs: 25 HIV negative mothers and 25 HIV positive mothers on long-term ART)
[c]The HIV positive group has a total of 31 individuals, including 22 males and 9 females, specific clinical group not specified
[d]The red box outlines the only two studies to date that investigated the relationship between gut microbiota and HIV in a sub-Saharan African cohort.
[e]Immunological responders (IR) and non-responders (INR) defined as having ≥350 and <350 CD4+ T-Cell counts/μL, respectively, after >2 years of ART.
Adapted from (Gootenberg et al., 2017). Data assembled from sources cited in **Table 1.5**.

Table 1.6 (Continued)

Study Design and Metadata																								
Study	McHardy et al.	Vujkovic-Cvijin et al.	Lozupone et al.[a,†]	Dillon et al.	Mutlu et al.	Yu et al.	Dinh et al.	Nowak et al.	Castellanos et al.	Dillon et al.	Dubourg et al.[c,†]	Ling et al.	Monaco et al.[d]	Noguera-Julian et al.[†]	Sun et al.	Yang et al.	Dillon et al.	Nowak et al.[a]
Pubmed ID	24451087	23843452	24034618, 25078714	24399150	24586144	24335481	25057045	26355675	25407519	25921339	27547442	27477587	26962942	27077120	27048741	26731752	28002063	28118207
Location	Los Angeles, CA	San Francisco, CA	Denver, CO	Denver, CO	Chicago, IL	Washington, DC and New York City, NY	Boston, MA	Stockholm, Sweden	Madrid, Spain	Denver, CO	Marseille, France	Hangzhou, China	Mbarara, Uganda	Barcelona, Spain and Stockholm, Sweden	Shanghai, China	New York City, NY	Aurora, CO	Abuja, Nigeria
Sample type	Rectal sponges, anal washes	Colon	Stool	Stool, fecal aspirates, colon	Stool, colon, ileum	Anal swabs	Stool	Stool	Stool	Rectal swabs, colon	Stool	Stool	Stool	Stool	Stool	Oral swab, esophogus, stomach, duodenum	Stool, colon	Rectal swabs
Region of 16S rDNA gene sequenced	V4	V1-V8	V4	V4	*Unspecified*	V3-V4	V3-V5	V3-V4	V1-V3	V4	V3-V4	V1-V3	V4	V3-V4	V3-V4	V3-V4	V4	V3-V4
16S sequencing method	Illumina	PhyloChip	Illumina	Illumina	454	Illumina	454	Illumina	454	Illumina	Illumina	454	Illumina	Illumina	454	454	Illumina	Illumina
Total number of participants	60	32	55	32	43	76	37	40	30	38	58	83	122	240	17	16	32	130
HIV uninfected (M/F)	20/0	9/0	15	9/5	17/5	51/0	12/4	5/4	8/7	9/5	27	16/0	20/20	28/6	2/2	4/4	9/5	55/0
Chronic infected, untreated (M/F)	20/0	6/0	12	13/5	-		-	3/6	-	18/6	13	32/0	11/31	100	1/1	5/3	13/5	41/0
Chronic infected, ART>12mo (M/F)	20/0	16/0	17	-	16/5	25/0	17/4	11/8	12/3	-	18	35/0	20/20	71	8/3	-	-	34/0
Chronic infected, ART<12mo (M/F)	-	-	11	-	(unspecified)	(unspecified)	-	(unspecified)	-	-	(unspecified)	-	-	13	-	-	-	(unspecified)
Acute infected (M/F)	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-
Long-term non-progressor M/F)	-	1/0	-	-	-		-	2/1	-	-	-	-	-	22	-	-	-	-
*Serum CD4+ T cells of HIV+ group (cells/μL)**	Untreated: 439±271; Treated: 534±246	Untreated: 356 (313-819); Treated: 374 (251-1110)	Untreated: 551±218; Treated: 483±258	425 (238-782)	334 (106-948)	Swab 1: 580 (432-721 IQR); Swab 2: 232 (32-415 IQR)	668 (424-870 IQR)	Before treatment: 355 (120-2470)	584 (466-794 IQR)	445 (221-1248)	Untreated: 361±475; Treated: 462±348	Untreated: 353±180; Treated: 363±185	Untreated: 225 (113-382 IQR); Treated: 396 (283-490 IQR)	Barcelona: 700 (462-860 IQR); Stockholm: 480 (380-630 IQR)	Untreated: 383 (11-756 IQR); Treated: 158 (36-414 IQR)	327 (12-708)	424.5 (238-782)	*Not reported*

Study Design and Metadata																								
Study	Pinto-Cardoso et al.	Serrano-Villar et al.	Vesterbacka et al.	Villanueva-Millán et al.	Armstrong et al.	Hoel et al.	Lee et al.	Lu et al.	Neff et al.	Qing et al.	San-Juan-Vergara et al.	Vázquez-Castellanos et al.[†]	Zhou et al.	Liu et al.	Maurice et al.	Sessa et al.	Williams et al.	
Pubmed ID	28262770	28000678	28740260	28362071	30396369	29712976	30250162	30034377	29650491	30560340	29540734	29789624	29411528	30685386	30649054	30649062	30618262	
Location	Mexico City, Mexico	Madrid, Spain	Stockholm, Sweden	Logrono, Spain	Denver, CO	Copenhagen, Denmark	Kuala Lumpur, Malaysia	Beijing, China	Denver, CO	Chengdu, China	Barranquilla and Cartagena, Columbia	Madrid, Spain	Guangzhou, China	Denver, CO	London, United Kingdom	Rome, Italy	Chicago, IL	
Sample type	Stool	Stool	Stool	Stool	Stool	Stool	Rectal swabs	Stool	Stool	Stool	Stool	Stool	Stool	Stool	Stool	Stool	Stool	Rectal swabs
Region of 16S rDNA gene sequenced	V3-V4	V1-V3	V3-V4	V4	V4	V3-V4	V4	*Unspecified*	V4	V4	V4	*Unspecified*	V4	V3-V4	V1-V2	V3-V4	V3-V4	*Unspecified*
16S sequencing method	Illumina	454	Illumina	Illumina	Illumina	Illumina	Illumina	Illumina	Illumina	Illumina	Illumina	Illumina, 454	Illumina	Illumina	Illumina	Illumina	Illumina	*Unspecified*
Total number of participants	43	44	64	71	217	84	46	91	50	25	37	54	68	36	79	132	100	
HIV uninfected (M/F)	6/4	6/3	8/8	11/10	64/41	33/7	20/0	30/0	19/5	10/0	7/5	15	20/15	21/1	15/2	33/38	0/30	
Chronic infected, untreated (M/F)	-	11/1	16/16	4/1	41/4	-	-	26/0	11/0	-	15/10	12	14/5	-	61/1	-	0/11	
Chronic infected, ART>12mo (M/F)	29/4	21/2 (13/2 IR,	-	30/15	52/15	41/3 (ART >11mo)	26/0 (16/0 IR, -	35/0 (15/0 IR,	8/7	15/0	-	27 (18 IR, 9	9/5 (greater than 3 mo)	14/0	(unspecified)	34/27	0/59	
Chronic infected, ART<12mo (M/F)	-	8/0 INR)[ε]	-	-	-	-	10/0 INR)[ε]	20/0 INR)[ε]	-	(unspecified)	-	INR)[ε]	-	-	-	-	-	
Acute infected (M/F)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Long-term non-progressor M/F)	-	-	9/7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
*Serum CD4+ T cells of HIV+ group (cells/μL)**	EFV ART: 457 (276-564.5 IQR); PI ART: 559 (369-703 IQR)	Untreated: 558 (432-646 IQR); INR: 291 (230-324 IQR); IR: 561 (426-794 IQR)	Elite controllers: 806 (676-1049 IQR); ART naïve: 390 (298-475 IQR)	Untreated: 40% 200-500, 60% >500; Treated: 27% 200-500, 73% >500	ART+ MSW: 699±311; ART- MSW: 962±344; ART+ F: 816±336; ART- MSW: 560±206; ART- MSW: 551±0; ART- F: 600±357	Without T2DM: 612 (490–734); With T2DM: 646 (511–780)	IR: 726 (656-1031 IQR); INR: 272 (262-338 IQR)	Untreated: 351.0 (182.8-432.8 IQR); Before treatment: 295.5 (214.0-352.0 IQR)	Treated MSW: 396 (177-1200); Untreated MSW: 478 (322-766); ART+ F: 735 (187-1417)	47% >500, 53% 200-500	Untreated: 528.9 (39.3–1,012.4)	*Not reported*	Untreated: 55±69; Treated: 172±136	570 (364 IQR)	With NAFLD: 815.5±309.2; Without NAFLD: 765.7±235.1	Prenatally HIV-infected: 829 (196-2172)	*Not reported*	

While there are overarching patterns that have emerged from these studies of HIV-associated microbial changes, a meta-analysis of a subset of these studies showed conflicting results and preliminarily concluded that no consensus could be reached on which bacterial genera were associated with HIV (Duvallet et al., 2017). Coupled with this is a need to more finely examine the taxa differentially abundant in HIV infection, since meaningful functional variation can often occur at the species or strain level rather than the family or genera level (Ahern et al., 2014; Faith et al., 2015; Ley, 2016). 16S rRNA gene sequencing cannot reliably provide this taxonomic resolution, creating a role for techniques such as shotgun metagenomic sequencing, which could detect strain or gene content variation, or isolation and full genome sequencing of microbes of interest.

1.5 Potential role of HIV-associated gut microbial changes in HIV disease progression

HIV-associated gut microbiome changes are frequently linked to systemic host inflammation and immune activation measured by circulating levels of the monocyte/macrophage membrane protein sCD14 as well as the marker of enteric epithelial disruption fatty-acid binding protein 2 (FABP2). CD14 functions as an LPS co-receptor (Kitchens, 2000; Ranoa et al., 2013; Tapping and Tobias, 2000) by forming a complex with LPS-binding protein, Toll-like receptor (TLR) 4, and MD-2 (Thomas et al., 2002; Yu and Wright, 1995) and can be shed as a soluble form (sCD14) that can act as a marker of intestinal barrier dysfunction and bacterial translocation (Brenchley and Douek, 2012; Brenchley et al., 2006; Jiang et al., 2009).

In light of these data and role of the enteric microbiome in many inflammation-associated pathologies such as diabetes mellitus, obesity, and IBD (Cadwell et al., 2010; Cho and Blaser, 2012; Friedrich, 2013; Gevers et al., 2014; Kostic et al., 2015;

Kostic et al., 2014; Le Chatelier et al., 2013; Littman and Pamer, 2011; Mazmanian et al., 2008; Peterson and Cardona, 2010; Peterson and Turnbaugh, 2010; Scher et al., 2015; Sze and Schloss, 2016; Willing et al., 2010), it has been proposed that HIV-associated microbiome shifts could contribute to inflammation-related NCD that are responsible for a large share of the increased mortality observed during chronic HIV infection (Centers for Disease Control and Prevention, 2008; Egger et al., 2002; Guaraldi et al., 2011; Lewden et al., 2007; Lohse et al., 2007; Reid et al., 2012).

One leading model is that the combined changes in the enteric immune system and microbial community could disrupt the balance of metabolic functions performed by the microbiota, such as short-chain fatty acid or bile acid metabolism (Brenchley and Douek, 2012a), or cause increased translocation of bacterial products into the systemic circulation (Brenchley et al., 2006). These insults to microbial metabolism and translocation of harmful bacterial products could lead to epithelial damage and chronic systemic inflammation (Brenchley and Douek, 2012; Brenchley et al., 2006; Dinh et al., 2015; Douek et al., 2009; Jiang et al., 2009; Reus et al., 2013; Sandler and Douek, 2012). With regard to metabolic imbalances, elevated plasma kynurenine, a tryptophan metabolite, has been found to be associated with CD8+ T-cell activation (Jenabian et al., 2015) and mortality (Hunt et al., 2014) in HIV-infected individuals. Additionally, HIV infection was associated with the presence of a gut microbial community with the genetic capacity to metabolize tryptophan into kynurenine and demonstrated kynurenine production *in vitro* (Vujkovic-Cvijin et al., 2013).

In relation to immune dysfunction, certain lymphocyte subsets are essential to the maintenance of enteric epithelial integrity (e.g. IL-22-producing ILC3 and Th17 cells

(Maloy and Kullberg, 2008; Thaiss et al., 2016)), so their destruction by HIV and the subsequent loss of immune surveillance may result in impaired intestinal epithelial barrier function, increased gut permeability, and further translocation of gut bacterial products into systemic circulation (Brenchley et al., 2006; Sandler and Douek, 2012). Translocated bacterial products could trigger persistent systemic immune activation and drive turnover of CD4+ and CD8+ T-cells leading to clonal exhaustion and progressive impairment of T-cell function (Torres et al., 2014). HIV-associated chronic systemic immune activation, secondary to a loss of gut epithelial homeostasis, may then lead to disease progression in both treated and untreated HIV infection. This model is supported by data showing that circulating levels of LPS, a bacterial product known to elicit an innate immune response by binding host TLRs (Cani et al., 2007), increase significantly as HIV infection progresses (Marchetti et al., 2013), though initiation of ART lowers circulating LPS levels (Brenchley et al., 2006). Individuals with both treated and untreated HIV infection, however, exhibit elevated serum levels of bacterial LPS and systemic inflammation relative to uninfected controls (**Figure 1.1**) (Brenchley et al., 2006; Cassol et al., 2010; Sandler et al., 2011). These phenomena may contribute to or synergize with the gastrointestinal (GI) sequelae that individuals experience when infected with HIV, including an enteropathy characterized by increased inflammation, diarrhea, and malabsorption (Brenchley, 2013).

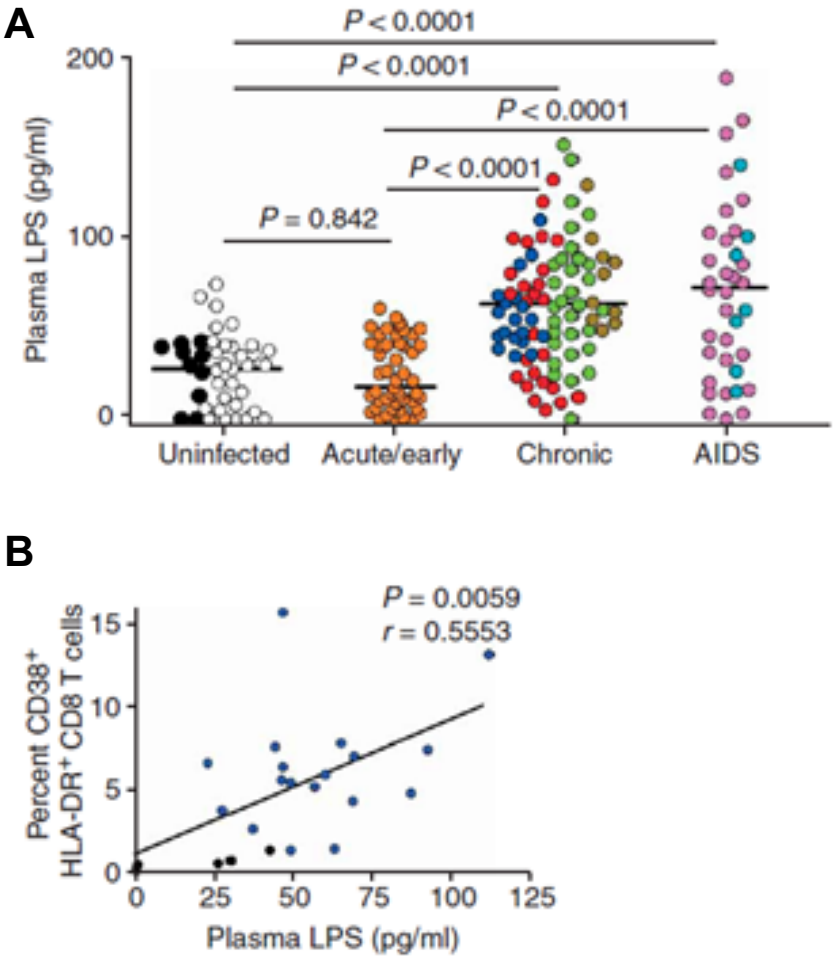

Figure 1.1: Plasma LPS and associated CD8+ T cell activation increase with HIV disease severity. (**A**) Plasma levels of LPS are increased significantly in chronic HIV and further increase in progression to AIDS. (**B**) Increasing plasma LPS correlates with increased CD8+ T cell activation as measured by CD38 and HLA-DR positivity. Adapted from (Brenchley et al., 2006).

In an analogous paradigm of metabolic dysfunction independent of HIV, the enteric microbiome has been shown to drive CVD pathogenesis by transforming dietary constituents such as phosphatidylcholine and bile acids into reactive intermediates such as trimethylamine N-oxide (TMAO) that can lead to macrophage and platelet activation, thrombosis, and arterial plaque formation (**Figure 1.2**) (Brown and Hazen, 2015; Koeth et al., 2014; Koeth et al., 2013; Wang et al., 2011; Zhu et al., 2016). Similarly, more direct microbiota-immune interactions have been associated with CVD in HIV-uninfected individuals, as microbiota-induced immune activation and inflammation has been shown to precipitate CVD pathogenesis (**Figure 1.2**) (Fisher et al., 2012; Khovidhunkit et al., 2004; Neves et al., 2013; Yuan et al., 2013). The HIV-associated findings of metabolic disruption such as kynurenine production and LPS-related immune activation suggest that pathology leading to CVD like that described above could occur in HIV as well, contributing to increased chronic mortality.

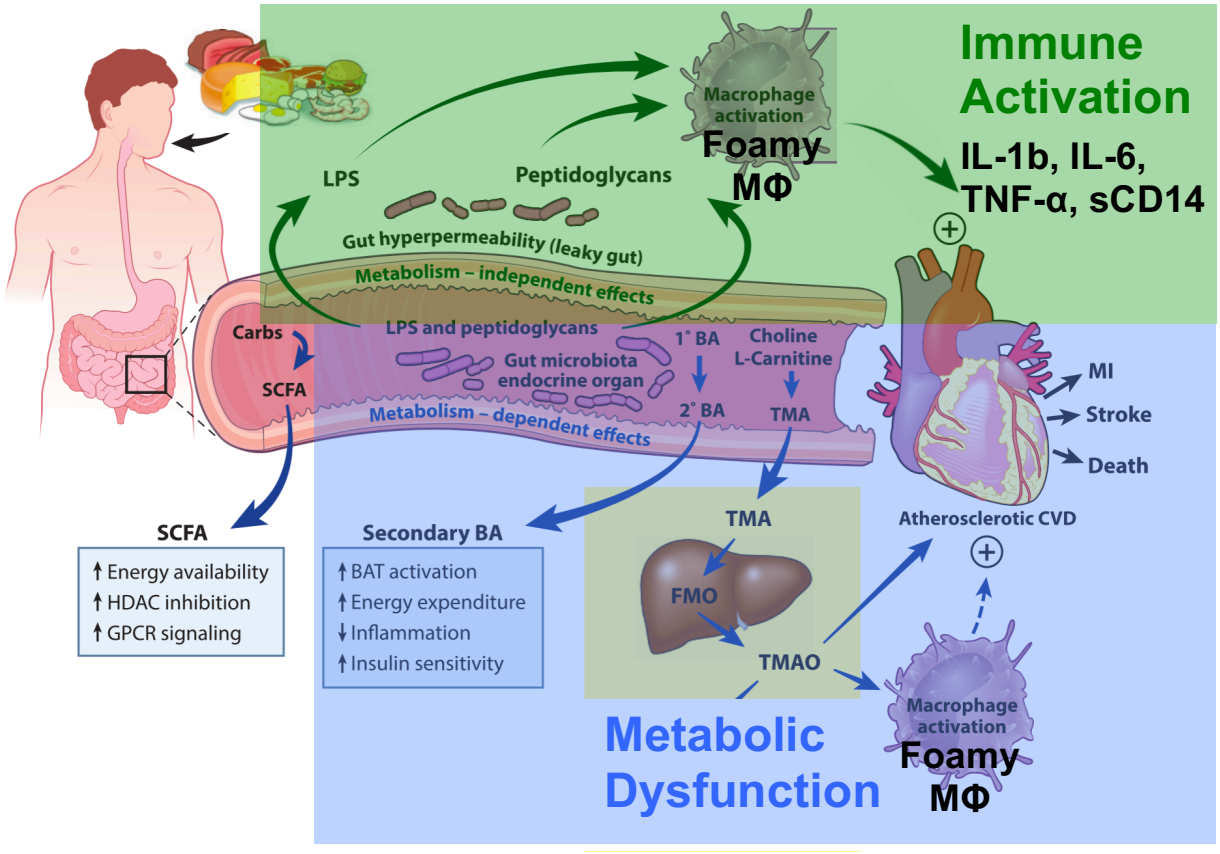

Figure 1.2: Models of enteric microbial induction of inflammation and CVD via immune or metabolic mechanisms. The microbiota could contribute to systemic inflammation and CVD by direct immune (green) or metabolic dysfunction (blue) mechanisms. In the direct immune mechanism, immune cells are activated by inflammatory bacterial products that translocate into systemic circulation. In the metabolic dysfunction mechanism, inflammatory molecules like TMAO are produced and this is coupled with the loss of commensal or anti-inflammatory function such as the production of SCFA. TMA: trimethylamine, TMAO: trimethylamine N-oxide, FMO: flavin-containing monooxygenase, SCFA: short-chain fatty acid, BA: bile acid (1˚: primary, 2˚: secondary), BAT: brown adipose tissue, HDAC: histone deacetylase, GPCR: G-coupled protein receptor. Adapted from (Brown and Hazen, 2015).

Because the microbiota has been demonstrated to contribute to host inflammation in both HIV-infected and -uninfected individuals through many mechanisms, further investigation of microbial changes associated with HIV infection has the potential to aid in the development of therapeutic interventions that could improve many of the previously mentioned pathologic inflammation-associated consequences of chronic HIV infection.

1.6 Geographical context for HIV-associated bacterial microbiota differences

Most current work investigating HIV and the enteric microbiome has focused on populations in industrialized high-income countries (HIC) (**Figure 1.3** and **Table 1.6**), as opposed to non-industrialized low-income countries (LIC) where HIV burden is greatest (UNAIDS, 2018). Since the burden of NCD in chronic HIV, most notably CVD, is growing rapidly in the HIV-positive population in sub-Saharan Africa (Hirschhorn et al., 2012; Hontelez et al., 2012; Justice and Braithwaite, 2012; Mills et al., 2012), this region could potentially benefit from the deployment of microbiota-directed therapeutics. However, there could be substantial obstacles to translating results from patients in HICs to those in LICs, as some of the greatest interpersonal variation in the community structure of human gut microbial communities is found between individuals in HICs and LICs (**Figure 1.4A**) (Smits et al., 2017).

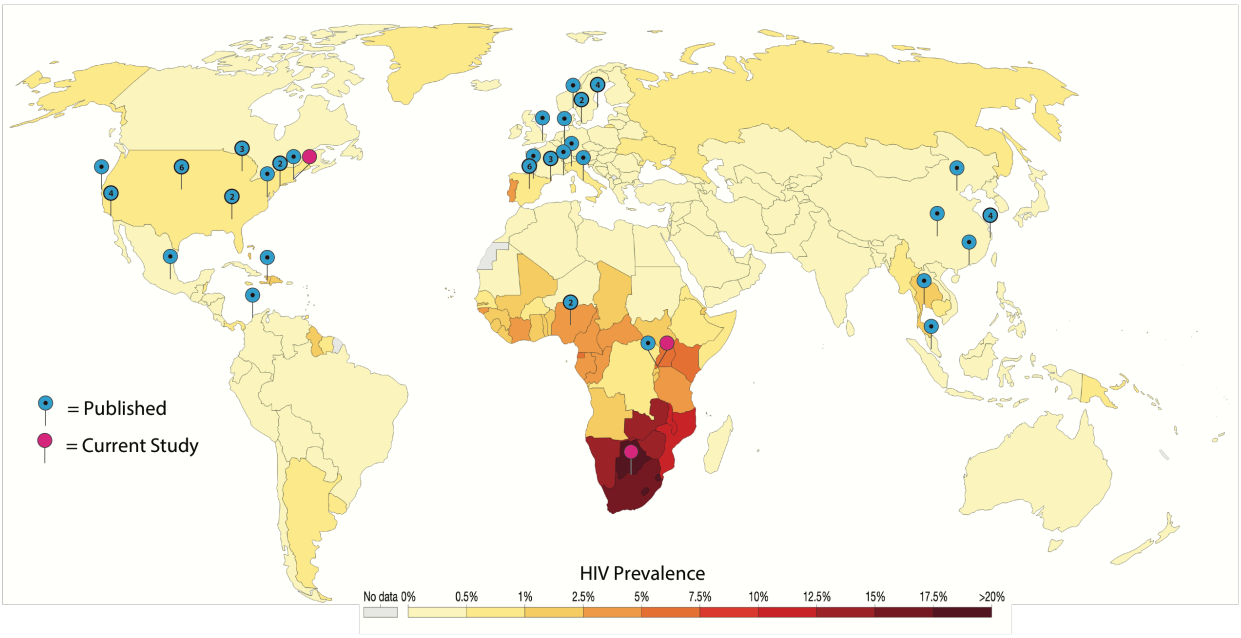

Figure 1.3: Locations of studies of HIV and the enteric microbiome in adults. A large majority of HIV-enteric microbiome studies have been conducted in HICs that are not the hardest hit by the HIV epidemic. Countries are shaded by HIV prevalence. Blue pins indicate locations of currently published studies. Locations with multiple studies have the number of studies indicated within the pin. Further data is available in **Tables 1.5** and **1.6**. Pink pins indicate the three cohorts discussed in Chapter 3 of this dissertation. Adapted from (Gootenberg et al., 2017) using data and map from (Global Burden of Disease Collaborative Network, 2018; Roser and Ritchie, 2019).

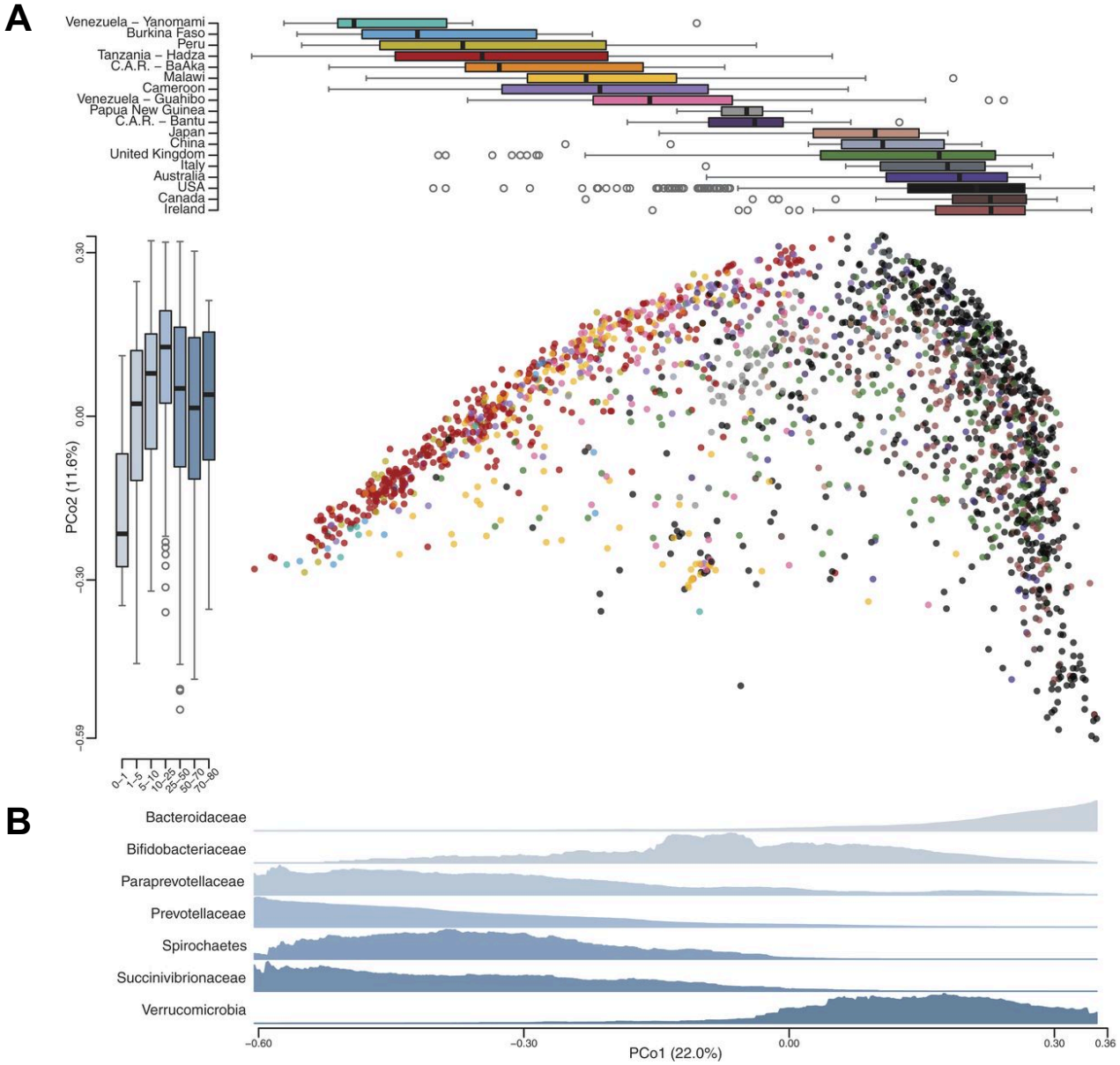

Figure 1.4: Meta-analysis of gut microbiome studies conducted across a wide range of geographies. (**A**) Bray-Curtis dissimilarity PCoA (center panel) based on 2064 microbial community compositions described at the family taxonomic level across populations. Each circle represents the placement of a microbial community projected in a subspace that maximizes the variance of the underlying taxonomic data; colors correspond to populations in the top panel. Boxplots (top panel) indicate the distribution of each population along the first principal coordinate (PCo1). Boxplots (left panel) depict the distribution of age (in years) according to gut microbial community placement on PCo2. Geography, as measured by study location, correlated highly with PCo1, the axis of greatest inter-sample variation. More rural and less-industrialized populations concentrated in LICs cluster on the left side of the figure while more urban and industrialized populations in HICs cluster on the right side of the figure. (**B**) Density plots of the moving average of seven taxa along PCo1, with a scale from zero to the maximum moving average. Adapted from (Smits et al., 2017).

The overall microbiota differences observed between HIV-infected individuals and uninfected individuals (Lozupone et al., 2013; Lozupone et al., 2014), especially the observations that HIV-infected individuals have increased abundances of organisms within the genus *Prevotella*, mirror some differences seen at baseline between populations in HICs and LICs. The gut microbiomes of subjects in HICs such as the United States or Western European countries are frequently characterized by higher abundances of organisms in the genus *Bacteroides, Bifidobacterium, and Ruminococcus* while those in LICs in Africa or South America have higher abundances of organisms in the genus *Prevotella*, *Succinivibrio*, and *Treponema* (**Figure 1.4B**) (De Filippo et al., 2010; Gomez et al., 2016; Gorvitovskaia et al., 2016; Lozupone et al., 2013; Martinez et al., 2015; Obregon-Tito et al., 2015; Schnorr et al., 2014; Smits et al., 2017; Wu et al., 2011; Yatsunenko et al., 2012). It is currently thought that the differences between HIC and LIC microbiota are highly related to the corresponding dietary differences, with greater consumption of fat and simple carbohydrates in HICs and greater consumption of more complex carbohydrates in LICs (David et al., 2014; Kovatcheva-Datchary et al., 2015; Wu et al., 2011). Recent findings indicate that an individual's degree of urbanization may contribute greatly to these geographical differences in the microbiome (Ayeni et al., 2018; Winglee et al., 2017). The observations that a Prevotellaceae-rich community is frequently observed in healthy individuals in LICs such as Burkina Faso, Venezuela, Malawi, or Papua New Guinea argue against the simple conclusion that defined HIV-associated taxa changes in the gut microbiota (e.g. an increase in Prevotellaceae and a decrease in Firmicutes) alone are responsible for chronic inflammation and pathology in HIV-infected individuals.

Rather, these potentially conflicting observations suggest a more complex relationship wherein a mismatch between the extant taxa and their host context causes inflammation (Lozupone et al., 2014). One possible example of this mismatch might be a microbial community that is inappropriate for the host diet and as a result produces inflammatory metabolites from these dietary constituents. This “context-dependent” model of microbial-driven systemic inflammation would further heighten the need to study HIV-associated microbial differences in appropriate populations, as therapeutic interventions constructed from data originating from one population may be poorly suited for application in a population with a different microbial and environmental milieu.

1.7 MSM in HIV and microbiota associations

In industrialized countries, men who have sex with men (MSM) are overrepresented in the population of people living with HIV. In the United States, MSM account for more than half of the estimated 1.1 million people living with diagnosed and undiagnosed HIV and comprise the majority of new HIV diagnoses (Centers for Disease Control and Prevention, 2019). On a more global scale, MSM sexual contact is the most significant risk factor for HIV acquisition in many HIC especially the United States and Western Europe (Fettig et al., 2014).

In this context, MSM as a population are disproportionately affected by HIV infection and could greatly benefit from microbiota-targeted strategies that could alleviate chronic symptoms and improve survival in HIV. However, analogous to how individuals in LICs have a significantly different baseline microbiota from most “healthy controls” in HICs, there is a growing body of evidence that MSM have a characteristic enteric microbial community. While there are numerous microbial differences that set

the gut microbiome of MSM apart from that of “healthy controls” in HICs, one of the most defining factors is the relative enrichment of bacteria within the Prevotellaceae family (Armstrong et al., 2018; Fulcher et al., 2018; Kehrmann et al., 2019; Kelley et al., 2017; Noguera-Julian et al., 2016; Nowak et al., 2019; Nowak et al., 2017; Pescatore et al., 2018). This difference is interestingly similar to those that exist between individuals living in LICs and HICs.

Investigation into the mechanism underlying these MSM-specific microbial communities has begun to produce data that provides insight into this phenomenon. Hyperosmolar sexual lubricants, which are frequently used during receptive anal intercourse and contain numerous solutes such as salts and carbohydrates, have been shown in a human trial to alter the rectal microbiome (Haaland et al., 2018). Two observational studies of sexual behavior and the microbiome found that condomless receptive anal intercourse in MSM was associated with significant changes in the rectal microbiome (Kelley et al., 2017; Pescatore et al., 2018). Two studies that probed sexual behavior and the microbiome proposed that sexual practices might underlie MSM-associated differences, though the specific sexual practices were not delineated in finer detail (Fulcher et al., 2018; Kenyon and Osbak, 2014). Another study that characterized MSM-associated microbial differences found that MSM reported a diet with more lean red meat, fewer servings of fruit, and less dietary fiber, possibly suggesting that diet could contribute to the observed enteric microbiota differences, though diet was not independently associated with major microbiota differences in linear modeling (Armstrong et al., 2018). While there is still much work to be done examining the mechanism accounting for MSM-characteristic gut microbial communities, the proposed

factors of sexual lubricant use, receptive anal intercourse, and diet are promising preliminary candidates for further investigation.

The prevailing observation that the gut microbiome of MSM is enriched for *Prevotella* means that this community is quite similar to that commonly found in individuals living in LICs and to that preliminarily proposed to be associated with HIV infection. Because MSM are significantly enriched within people living with HIV in HICs, the observed microbial community similarities may precipitate confounding of these associations and could be contributing to the varied findings and disagreements among studies of HIV and the gut microbiome (Duvallet et al., 2017; Noguera-Julian et al., 2016). Therefore, analogous to the need to characterize the gut microbiome of both HIV-infected and HIV-uninfected individuals in sub-Saharan African countries, there is an imperative to probe the gut microbial communities of both HIV-infected and HIV-uninfected MSM so as to better target this research towards producing findings that will be more applicable to those populations of greatest need in the context of the global HIV epidemic.

1.8 HIV-associated microbiome differences and body sampling site

Some of the variation in the findings among HIV-microbiome studies may be attributed to the differences in body site sampled and sampling methodology. The predominant collection method was stool sampling (**Table 1.6**), which may be most representative of the luminal microbial community, though anal swabs and washings or mucosal biopsies were also used. Independent of HIV, mucosal and luminal microbial communities have been shown to differ by varying degrees (Bajaj et al., 2012; Eckburg et al., 2005; Morgan et al., 2012; Stearns et al., 2011; Yasuda et al., 2015; Zoetendal et

al., 2002), though in some contexts, the mucosal and luminal communities correlate strongly and are representative of one another (Yasuda et al., 2015). With regard to functional differences, facultative anaerobes have been found to be more abundant in mucosal-associated environments, while obligate anaerobes are more prevalent in the gut lumen (Yasuda et al., 2015). Studies of the HIV-associated microbiota that used both mucosal- and luminal-targeted sampling techniques (Dillon et al., 2014; Dillon et al., 2016; Mutlu et al., 2014) found a variety of differences between these techniques. In the most concordant finding, the two sampling sites produced similar conclusions, with most taxa showing the same patterns of HIV-associated enrichment or depletion at both sampling sites (**Tables 1.5** and **1.6**) (Mutlu et al., 2014). However, the HIV-associated reduction in alpha diversity was more pronounced in mucosa than stool samples. Other studies found greater variation in the HIV-associated differences by body site. For instance, Dillon et al. 2016 detected an HIV-associated increase in abundance of Prevotellaceae at both sampling sites, but only observed HIV-associated decreases in taxa in the phyla Bacteroidetes and Firmicutes and increases in Proteobacteria in the mucosal samples (Dillon et al., 2016). The study with the greatest discordance between mucosal and luminal findings observed many more HIV-associated differences in mucosal-associated communities than luminal communities (Dillon et al., 2014). In addition, mucosal community composition changes were more closely associated with mucosal cellular immune activation. These results suggest that mucosal findings may be more sensitive or representative of the causative community, possibly because these microorganisms are in the closest contact with epithelial cells and immune cells. These advantages, however, must be weighed against the relative difficulty of obtaining

mucosal samples, especially in resource-limited settings or large cohort studies. Greater coordination of methodologies would allow more robust analysis of multiple studies and potentially reconcile conflicting conclusions.

1.9 Examining HIV-associated microbiota differences in different host contexts

The role of enteric microbial changes in HIV disease progression has been the focus of increasing investigation. Models regarding the connections between gut microbiome changes and chronic HIV pathogenesis hypothesize a role for gut epithelial damage and systemic immune activation as an intermediate mechanism. Investigations into HIV-associated differences in the gut microbial community have found varied changes, but a few overall consistent patterns have emerged. In general, HIV infection in HICs is associated with decreases in many bacterial families within the phyla Bacteroidetes and Firmicutes as well as increases in the family Prevotellaceae and families within the phyla Proteobacteria and Fusobacteria. This community shift implies an overall pathogenic or pro-inflammatory outcome based on the functions of the differentially abundant microorganisms, but this causative relationship has not been conclusively shown and bacterial behavior can vary widely based on context.

The heterogeneity of the conclusions drawn to date within this field may be due in part to the variation in study populations and methods employed (**Table 1.6**). Clinical application of this research has struggled and microbiota-targeted probiotic interventions have produced mixed results, with some therapeutics reducing inflammatory markers and others lacking efficacy (Cunningham-Rundles et al., 2011; d'Ettorre et al., 2015; González-Hernández et al., 2012; Hummelen et al., 2011; Irvine et al., 2010; Miller et al., 2016; Monachese et al., 2011; Reid, 2010; Schunter et al., 2012;

Stiksrud et al., 2015; Villar-Garcia et al., 2015). In light of the current findings of HIV-associated Prevotellaceae enrichment and the great abundance of this bacterial family in LIC populations that constitute the majority of HIV-infected individuals, there is a dire need for further examination of HIV-associated gut microbial differences in non-industrialized populations. It is possible that therapeutic strategies that consider Prevotellaceae enrichment a pathogenic state would be ill suited for HIV-infected individuals in LIC. Further understanding of the enteric microbial changes associated with HIV infection, especially among LIC populations that bear the greatest burden of HIV infection, is therefore necessary to design therapeutic strategies that could alleviate the sequelae of systemic inflammation and NCD in chronic HIV infection.

To address these prior shortcomings, here we examine the effect of HIV on the enteric bacterial microbiome in three distinct geographical locations spanning two continents, representing regions that are particularly relevant to the current HIV epidemic. We demonstrate that there are unique HIV-associated gut microbial differences at each location, both at the whole-community and individual taxa levels. These microbial differences also coincide with differences in the relationship between microbial taxa and host inflammatory markers. Taken together, these findings suggest that HIV infection could alter anti-inflammatory microbial functions and allow the emergence of new functions, though all of these effects are highly context-dependent on the pre-existing gut microbial community. This model has significant implications for efforts that aim to translate the results of HIV-microbiome studies between different geographical locations.

Chapter 2: Investigation of HIV-associated Gut Microbiome Differences in Uganda

2.1 Need for investigation of HIV-gut microbiome association in sub-Saharan Africa

Despite the unequal burden of HIV in developing regions of sub-Saharan Africa (HIV/AIDS, 2013) and the substantial known differences in baseline gut microbiome between individuals in HICs and LICs (Smits et al., 2017), prior to 2016 no studies of HIV infection and gut microbiome composition had been undertaken in Africa (**Table 1.6**). In contrast, nine studies had been conducted in HICs, predominantly in North America and Europe. The overarching conclusions arising from these studies had begun to coalesce around a model in which bacteria from the genus *Prevotella* were increased in abundance in HIV-infected individuals relative to uninfected controls and that these *Prevotella* were acting in a pro-inflammatory role. However, this conclusion would seem paradoxical in light of the fact that *Prevotella* comprise a large fraction of the baseline microbiota of healthy individuals in lower-income countries and less-urbanized areas (De Filippo et al., 2010; Smits et al., 2017). This discordance between quantity of investigation and prevalence of HIV infection had the potential to lead to situations in which interventions derived from gut microbiome studies conducted in high-income countries might not be efficacious in low-income countries, where therapies to alleviate HIV infection are most desperately needed.

To remedy this oversight, we partnered with colleagues in the MGH Global Health Division to investigate the gut microbiota of HIV-infected and -uninfected individuals living in the rural region surrounding the city of Mbarara in Southwestern Uganda. The following data is in large part presented in (Monaco et al., 2016), which also covered the viral component of the gut microbiota (virome). I did not participate in the generation and analysis of data pertaining to virome investigation in this paper and as such those findings will not be included in this chapter.

Adapted from (Monaco et al., 2016)

2.2 Characteristics of the first Ugandan cohort

Under the auspices of the Uganda AIDS Rural Treatment Outcomes (UARTO) study, 40 HIV-infected subjects on antiretroviral therapy (ART; “HIV-treated”) and 42 HIV-infected, ART-untreated (“HIV-untreated”) subjects were recruited at the Immune Suppression Syndrome Clinic at Mbarara Regional Referral Hospital. A comparison group of 40 HIV-uninfected subjects (“HIV-uninfected”) was recruited from individuals who had recently received a negative HIV test result at the HIV testing clinic (**Figure 2.1** and **Table 2.1**).

Adapted from (Monaco et al., 2016)

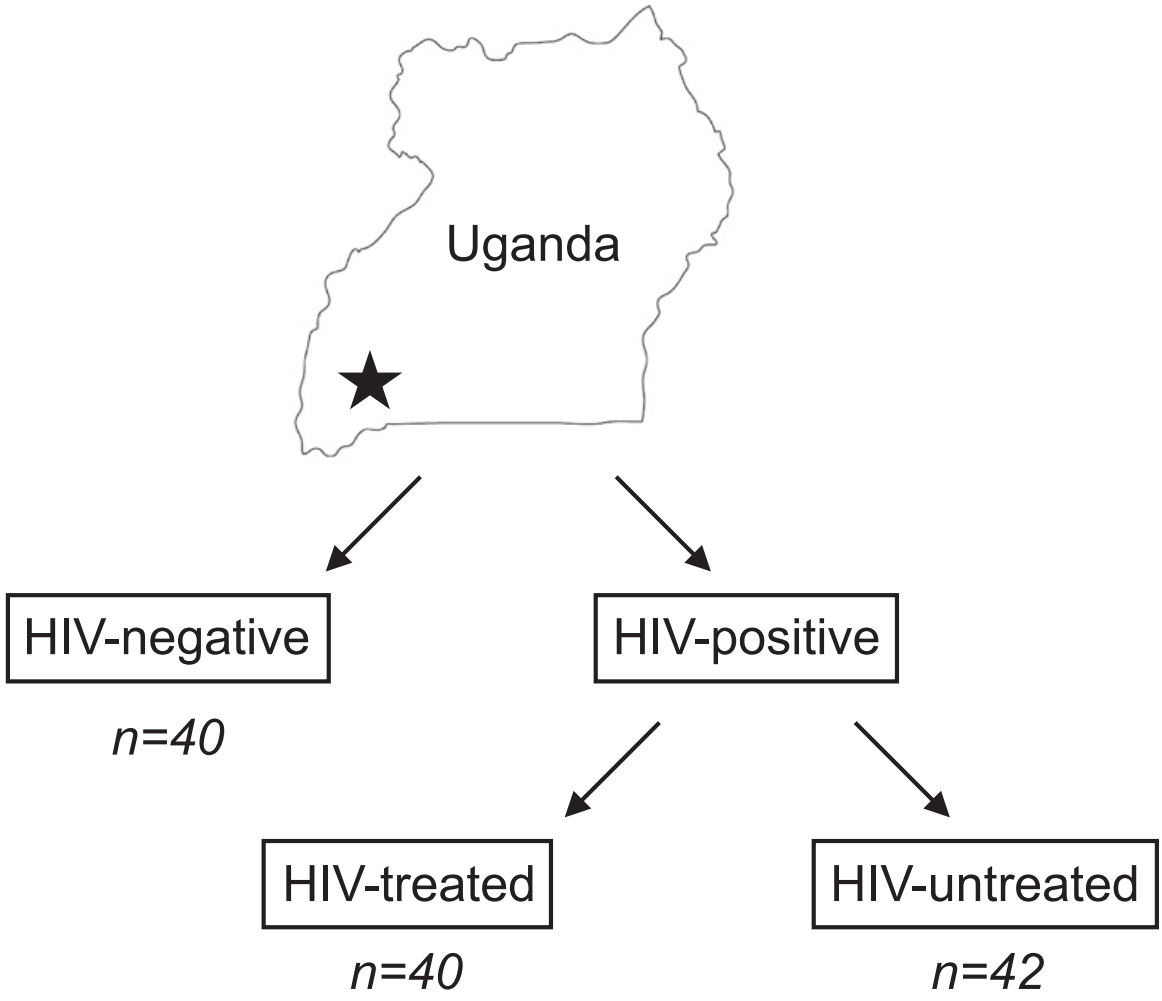

Figure 2.1: Study design and geographic location. The location of the city of Mbarara in southwestern Uganda is indicated with a star. The HIV and ART treatment statuses of the members of the cohort are shown below. Figure from (Monaco et al., 2016).

Adapted from (Monaco et al., 2016)

Among HIV-infected subjects who were receiving ART, the median time on ART was 6.7 years (interquartile range, IQR, 6.1-7.1). At the time of sampling, 35 of the 40 HIV-treated subjects (87.5%) had suppressive ART therapy as defined by undetectable viral loads (<20 viral RNA copies/mL). Out of the 82 total HIV-infected individuals, 57 (79.5%) had CD4+ T cell counts measured as greater than 200 cells/μl ("HIV CD4>200"), while 25 (20.5%) had CD4+ T cell counts measured as less than 200 cells/μl ("HIV CD4<200"), which indicates an immunocompromised host state. There were significant differences in the CD4+ T cell counts, CD4+ T cell percentages, HIV viral load, prophylactic co-trimoxazole use, age, sex and other variables between HIV-treated and untreated groups (**Table 2.1**). With respect to relationships between immune status and viral replication, peripheral CD4+ T cell counts in HIV-infected subjects were negatively correlated with circulating HIV viral load in subjects with detectable peripheral HIV RNA (**Figure 2.2A,** $p < 0.0001$). This would be expected, as greater viral replication would result in the destruction of host CD4+ T cells. With respect to associations between immune status and patient characteristics, peripheral CD4+ T cell count was positively correlated with weight ($p = 0.0037$) and body mass index (BMI, $p = 0.0093$) (**Figures 2.2B** and **2.2C**). This is consistent with the connection between HIV disease progression and overall health, as individuals with a more robust immune status as measured by CD4+ T cell count would have better overall health, as measured by weight. These clinical characteristics are consistent with the known effects of HIV infection (Mankal and Kotler, 2014; Palermo et al., 2011; Phillips et al., 2010).

Adapted from (Monaco et al., 2016)

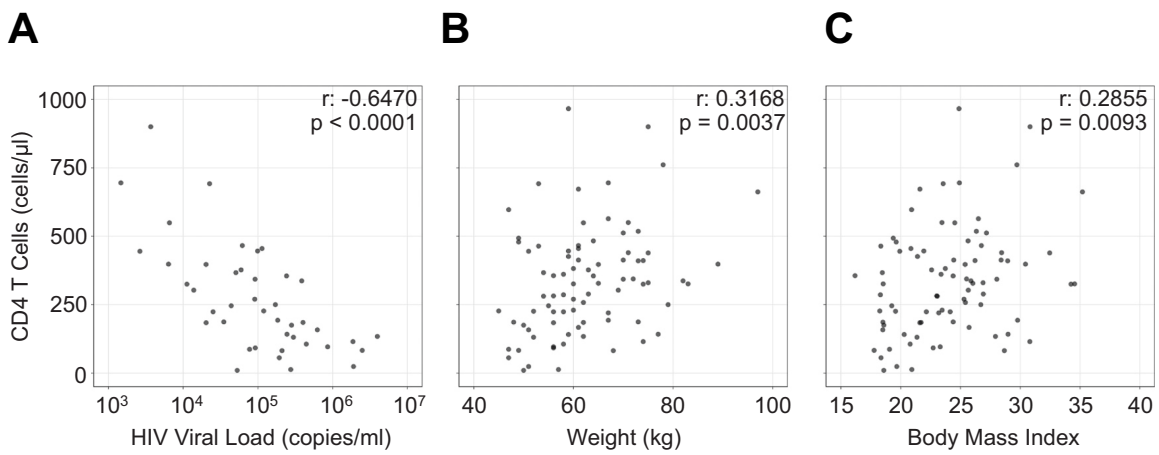

Figure 2.2: Factors correlated with CD4+ T cell count. Correlation between CD4+ T cell count (y-axis) compared to (A) HIV viral load ($p < 0.0001$); (B) weight ($p = 0.0037$); and (C) body mass index ($p = 0.0093$). Figure from (Monaco et al., 2016).

Adapted from (Monaco et al., 2016)

Table 2.1: First Ugandan Cohort Characteristics. NA, not available; NS, not significant; IQR, interquartile range; OI, opportunistic infection. For comparing continuous variables, Mann-Whitney and Kruskal-Wallis tests were used; for comparing categorical variables, Chi-Square and Fisher's exact tests were used. [a], NS between HIV-infected groups. Table from (Monaco et al., 2016).

Adapted from (Monaco et al., 2016)

Table 2.1 (Continued)

Patient Characteristics	HIV Neg (n = 40)	HIV on ART (n = 40)	HIV No Treatment (n = 42)	p Value
16S rRNA sequencing (n) (total 110)	37	39	34	
VLP NGS sequencing (n) (total 65)	21	21	23	
Demographics, median (IQR)				
Age	43 (38–48 years)	44 (38–49 years)	29 (24–34 years)	<0.0001
Male, n	20 (50%)	20 (50%)	11 (26.2%)	0.0404
Height	159 (155–163 cm)	159 (157–168 cm)	161 (156–166 cm)	0.3894
Weight	64 (55–74 kg)	61 (56–71 kg)	60 (53–68 kg)	0.5735
BMI	24 (22–28)	24 (21–27)	23 (20–27)	0.4305
Laboratory measures, median (IQR)				
CD4 T cell count	NA	396 (283–490 cells/μL)	225 (113–382 cells/μL)	0.0003
CD4 T cell count >500, n	NA	8 (20%)	4 (9.5%)	
CD4 T cell count 200–500, n	NA	29 (72.5%)	16 (38.1%)	
CD4 T cell count <200, n	NA	3 (7.5%)	22 (52.4%)	
CD4 Percent	NA	25 (21–31)	15 (10–24)	<0.0001
HIV Viral Load	NA	20 (20–20 copies/mL)	95,571(24,455–285,548 copies/mL)	<0.0001
CD4 nadir	NA	116 (58–167 cells/μl)	225 (110–382 cells/μL)	0.0001
Symptoms over last 30 days				
Nausea/Vomiting, n	13 (32.5%)	5 (12.5%)	15 (35.7%)	0.0389
Diarrhea, n	10 (25%)	4 (10%)	9 (21%)	0.2000
Constipation, n	17 (42.5%)	13 (32.5%)	10 (23.8%)	0.1969
Loss of appetite, n	21 (52.5%)	10 (25%)	20 (47.6%)	0.0286
Dysgeusia, n	20 (50%)	14 (35%)	17 (40.5%)	0.3875
Medications last 30 days				
ART, n	NA	40 (100%)	NA	
ART: NRTI, n	NA	40 (100%)	NA	
ART: NNRTI, n	NA	35 (85%)	NA	
ART: PI, n	NA	6 (15%)	NA	
Years on ART, median	NA	6.7 (6.1–7.1 IQR)	NA	
OI prophylaxis: bactrim last 30 days, n	0 (0%)	38 (95%)	38 (90.5%)	<0.0001[a]
Other Antimicrobials 3 days, n	8 (20%)	5 (12.5%)	4 (9.5%)	0.3721
Other Antimicrobials 30 days, n	15 (37.5%)	14 (35%)	4 (9.5%)	0.0066
Living conditions				
Water source, n				
Water source: communal tap	NA	15 (37.5%)	22 (52.4%)	0.1913
Water source: piped in	NA	2 (5%)	2 (4.8%)	1.0000
Water source: open well	NA	7 (17.5%)	8 (19%)	1.0000
Water source: protected well	NA	7 (17.5%)	1 (2.4%)	0.0274
Water source: other	NA	9 (22.5%)	9 (21.4%)	1.0000
Distance to water, median	NA	75 m (6–675 IQR)	15 m (6–350 IQR)	0.3071
Toilet type: covered pit latrine, n	NA	25 (62.5%)	14 (33.3%)	0.0145
Toilet type: uncovered pit latrine, n	NA	15 (37.5%)	25 (59.5%)	0.0512
Toilet type: flush toilet, n	NA	0	3 (7.1%)	0.2412
Food Security				
Grows own produce, n	NA	29 (72.5%)	20 (47.6%)	0.0260
Owns livestock, n	NA	19 (47.5%)	12 (28.6%)	0.1107
Goes hungry, n	NA	11 (27.5%)	12 (28.6%)	1.0000
Cooks in kitchen, n	NA	25 (62.5%)	18 (42.9%)	0.0828
District				
MBARARA, n	30 (75%)	26 (65%)	27 (64.2%)	0.5135
ISINGIRO, n	3 (7.5%)	6 (15%)	9 (21.4%)	0.2057
KIRUHURA, n	4 (10%)	2 (5%)	4 (9.5%)	0.6655
Other or NA, n	3 (7.5%)	6 (15%)	2 (4.8%)	0.2485

ART, antiretroviral therapy; NA, not available; NS, not significant; IQR, interquartile range; for comparing continuous variables, Mann-Whitney and Kruskal-Wallis tests were used; for comparing categorical variables, chi-square and Fisher's exact tests were used. OI, opportunistic infection.
[a]NS between HIV-positive groups.

Adapted from (Monaco et al., 2016)

2.3 Serum soluble CD14 levels are elevated in HIV-infected individuals and correlate with increased HIV viral load and loss of CD4+ T cells

Soluble CD14 (sCD14) is a co-receptor that serves to bind bacterial lipopolysaccharide (LPS) and high blood concentrations of the protein are used to indicate increased amounts of microbial translocation across the gut epithelium from the gut lumen to the underlying tissue and systemic circulation (Marchetti et al., 2013). HIV-infected individuals have been shown to have elevated levels of circulating sCD14, and the levels of this inflammatory marker correlate with patient mortality (Sandler et al., 2011). In our cohort, HIV-infected individuals (both those receiving ART treatment and those with untreated HIV infection) had significantly higher levels of plasma sCD14 than HIV-uninfected subjects (**Figure 2.3A**, $p = 0.0017$), indicating that they exhibited a greater extent of bacterial translocation associated inflammation. Plasma levels of sCD14 were further elevated in patient subsets with more severe disease, as HIV-infected subjects with peripheral CD4+ T cell counts less than 200 cells/μl, which is indicative of profound immunodeficiency and advanced HIV disease (Selik et al., 2014), had levels of circulating sCD14 that were significantly higher than both HIV-negative subjects and HIV-infected subjects with CD4 >200 (**Figure 2.3B**, $p < 0.0001$).

Plasma levels of sCD14 correlated with clinical measures of disease severity within these patient subsets as well, as plasma sCD14 concentrations were positively correlated with circulating HIV viral load in HIV CD4<200 patients (**Figure 2.3C,** $p = 0.002$). From these data, we can see that subjects in our cohort conform to the expected positive correlation between sCD14, a marker of disrupted intestinal function and systemic inflammation (Marchetti et al., 2013), and HIV-associated

Adapted from (Monaco et al., 2016)

immunodeficiency as measured by peripheral CD4+ T cell loss and elevated levels of circulating HIV virus.

Adapted from (Monaco et al., 2016)

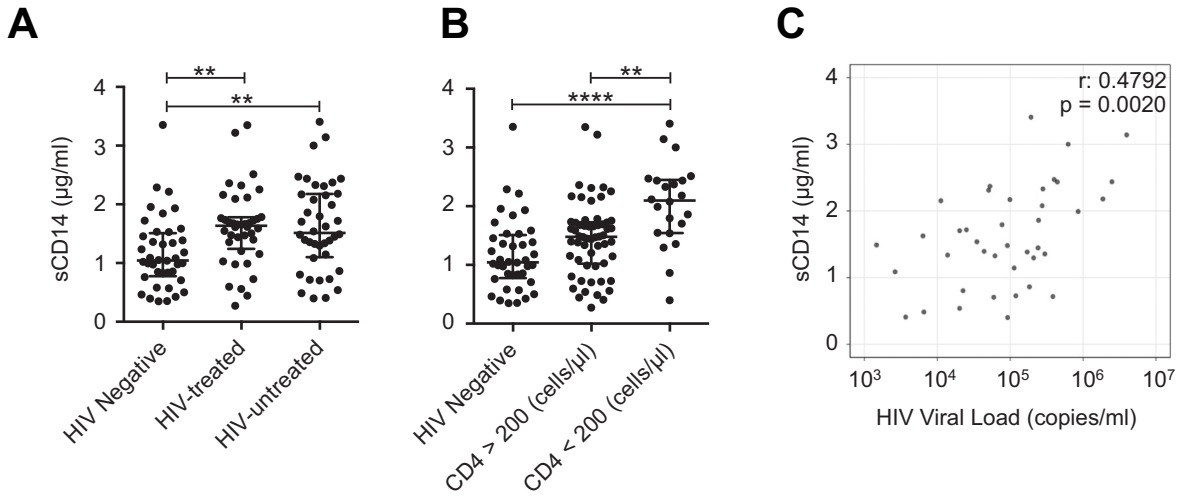

Figure 2.3. Plasma sCD14 is increased in HIV infection and correlates with HIV viral load. Circulating levels of sCD14 were graphed by (A) HIV status and ART treatment group and (B) CD4 T cell count. (c) Correlation between sCD14 levels and HIV viral load (p =0.002). p ≤ 0.01 = **, p ≤ 0.0001 = ****. Bars indicate median ± interquartile range (IQR). Figure from (Monaco et al., 2016).

Adapted from (Monaco et al., 2016)

2.4 Gut microbiome community characteristics are most correlated with peripheral CD4+ T cell counts and HIV status

We used amplicon sequencing of the 16S rRNA gene and computational analysis with Quantitative Insights Into Microbial Ecology (QIIME) (Caporaso et al., 2010a; Caporaso et al., 2010b; DeSantis et al., 2006; Edgar, 2010; Lozupone and Knight, 2005; McDonald et al., 2012; Price et al., 2010; Wang et al., 2007) and Phyloseq (McMurdie and Holmes, 2012) to characterize the enteric bacterial microbiome of 110 subjects (37 HIV-uninfected; 39 HIV-infected, ART-treated; 34 HIV-infected, ART-untreated) within our cohort (**Figure 2.4A, 2.5**). The most severely immunocompromised HIV-infected subjects, as indicated by a peripheral CD4+ T cell count less than 200 cells/μl had significantly decreased bacterial richness, as measured by the Chao1 index, compared to HIV-infected subjects with peripheral CD4+ T cell counts greater than 200 cells/μl (**Figure 2.4B,** $p = 0.009$), and the decrease in bacterial richness from HIV-uninfected subjects to HIV CD4<200 subjects was trending towards significance (**Figure 2.4B,** $p = 0.066$). Similarly, the bacterial phylogenetic diversity, as measured by Faith's Phylogenetic Diversity, was significantly lower in HIV CD4<200 subjects than both HIV CD4>200 subjects (**Figure 2.4C,** $p = 0.018$) and HIV-uninfected subjects (**Figure 2.4C,** $p = 0.036$). While levels of immunocompetence, as measured by peripheral CD4+ T cell counts, did correspond with differences in alpha diversity measures between groups of HIV-infected subjects, we did not detect a relationship between alpha diversity and treatment status, as there was no significant difference in bacterial richness or phylogenetic diversity between HIV treatment groups. Overall, peripheral CD4+ T cell count appeared to have the strongest relationship with intra-

Adapted from (Monaco et al., 2016)

sample bacterial diversity metrics, as HIV CD4<200 subjects had decreased richness and diversity, though HIV status was also related to these metrics, albeit less strongly.

To determine whether our factors of interest were associated with inter-microbial community differences, we calculated the inter-sample β-diversity as measured by weighted UniFrac distance (Lozupone and Knight, 2005). As tested with Permutational Multivariate Analysis of Variance Using Distance Matrices (PERMANOVA), there was no significant difference between patient grouping when subset by HIV infection status ($p = 0.588$) or peripheral CD4+ T cell count ($p = 0.231$). Similarly, principal coordinates analysis (PCoA) transformation of weighted UniFrac distances did not reveal any clustering among subjects when subset by peripheral CD4+ T cell count (**Figure 2.4D**) or HIV status (**Figure 2.4E**). Similar analysis using PCoA transformation of weighted UniFrac distances did not detect relationships between bacterial community structure and sex, water source, recent diarrhea, antibiotic usage, or home geographic district (**Figure 2.6**).

Adapted from (Monaco et al., 2016)

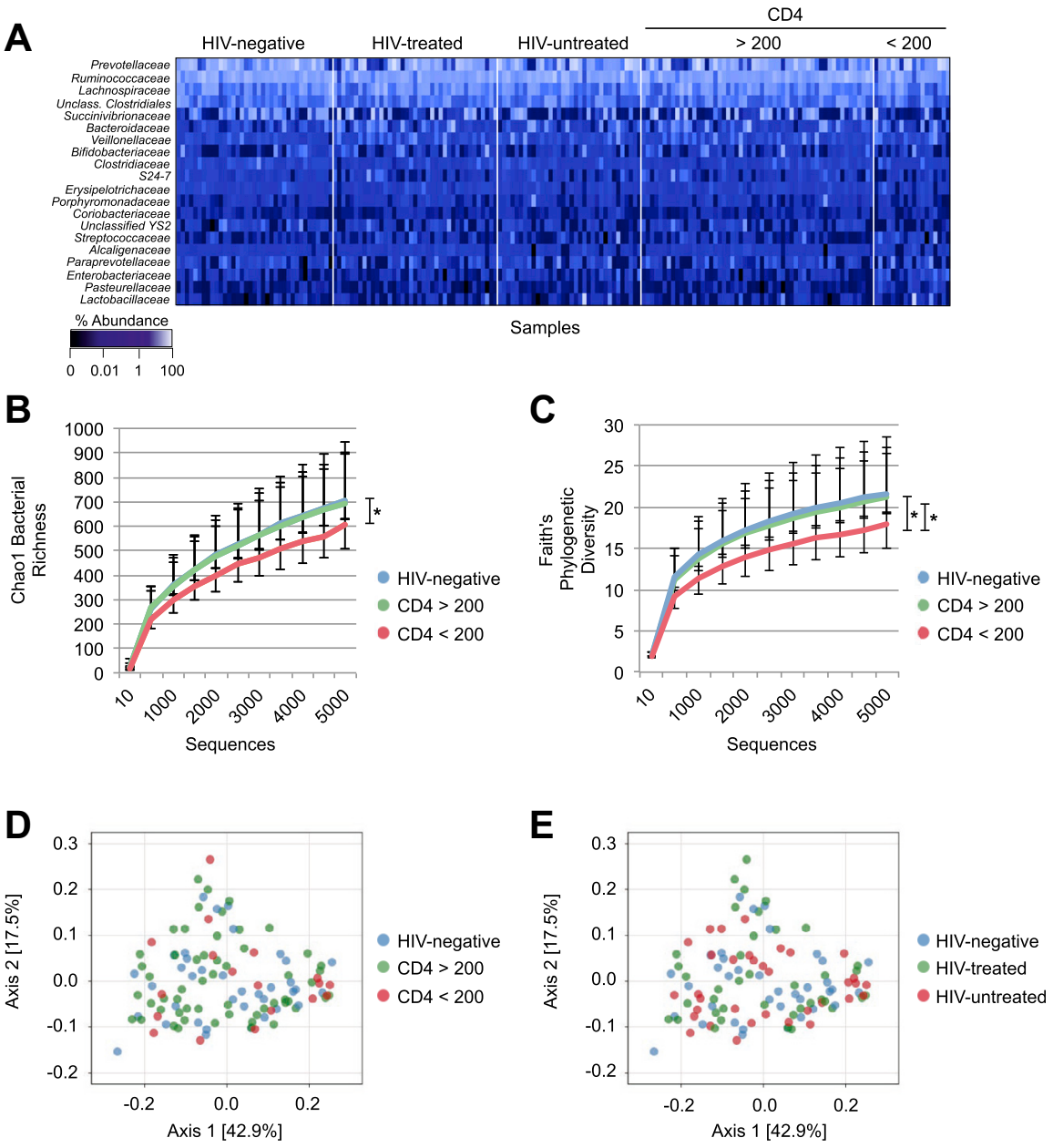

Figure 2.4. Bacterial community profiling. (A) Heatmap showing relative abundance of the 20 most frequent bacterial families (y-axis) by sample (x-axis), grouped by HIV status and CD4+ T cell count. Percent abundance is indicated by the gradient key. (B) Chao1 rarefied bacterial richness colored by CD4 T cell count. (C) Comparison of bacterial Faith's phylogenetic diversity in HIV-uninfected and HIV-infected subjects by CD4 T cell count. Statistical analysis was performed in QIIME using two-sample, non-parametric t-tests with Monte Carlo permutations. Error bars indicate SEM. $p \leq 0.05$ = *. Principal Coordinate Analysis (PCoA) plots of the weighted UniFrac distances colored by HIV status and (D) CD4+ T cell count or (E) ART treatment. Figure from (Monaco et al., 2016). **See also Figure 2.5, 2.6**.

Adapted from (Monaco et al., 2016)

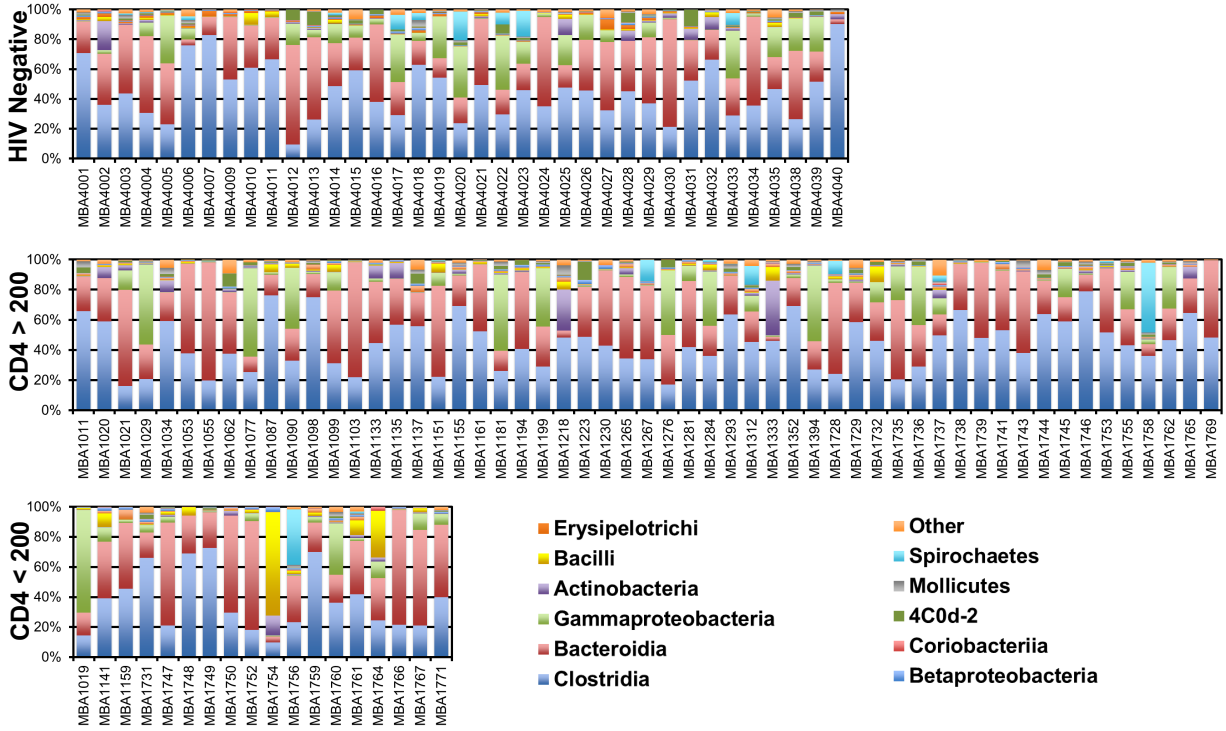

Figure 2.5: Bacterial community structure by relative abundance at class level split by HIV status and CD4+ T cell count. Relative abundance of bacterial taxa (y-axis) assigned to subjects (x-axis) as determined by 16S rRNA gene V4 amplicon sequencing was plotted and grouped by HIV-uninfected subjects (upper panel) and HIV-infected subjects with peripheral CD4+ T cell count above 200 cells/μl (middle panel) or below 200 cells/μl (lower panel). Taxonomic class color assignment key is located at bottom right. Figure from (Monaco et al., 2016).

Adapted from (Monaco et al., 2016)

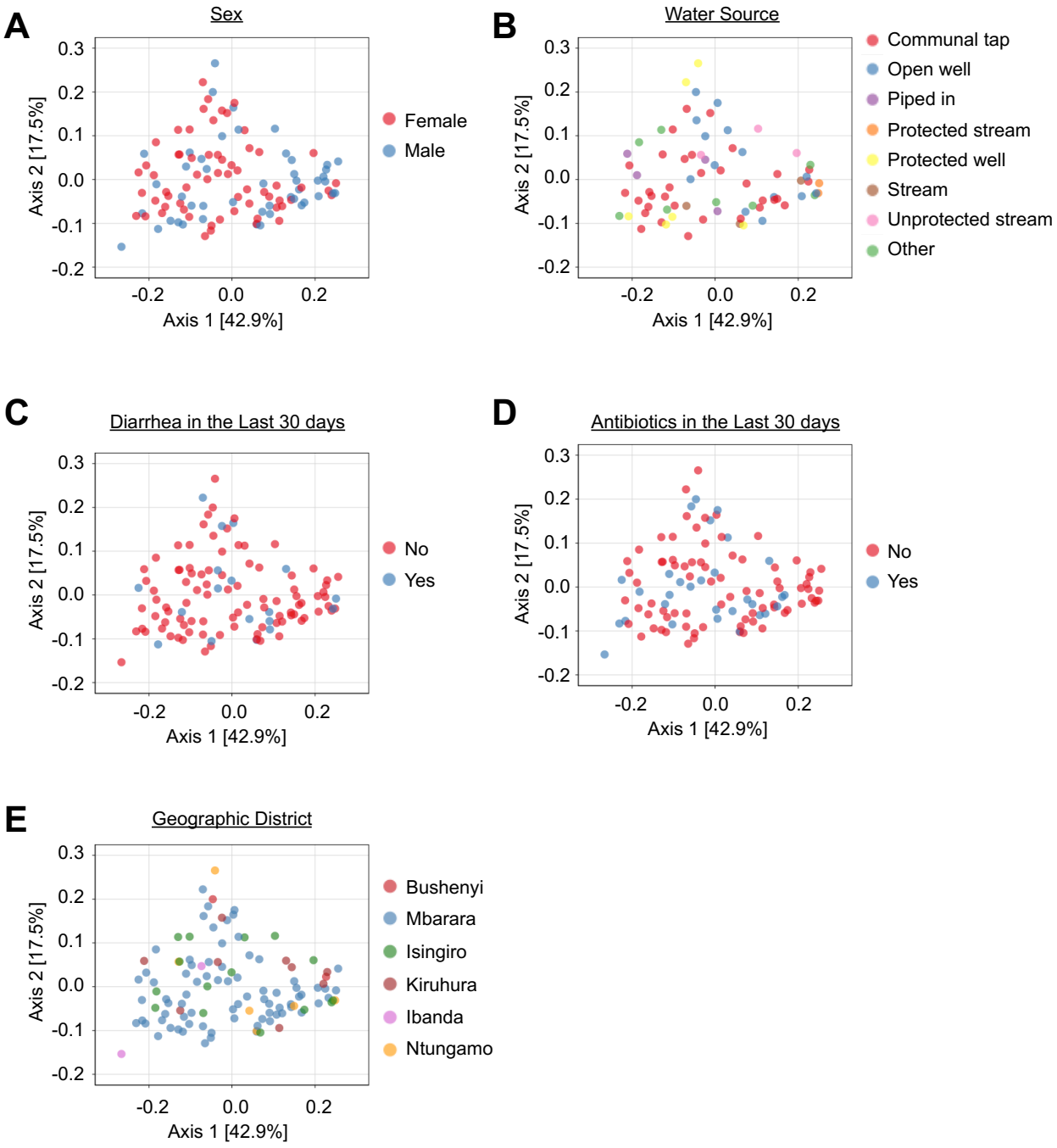

Figure 2.6: Environmental and clinical associations with bacterial community structure. Principal Coordinate Analysis (PCoA) plots of the weighted UniFrac distances colored by (A) gender, (B) patient's home water source, (C) reported diarrhea or (D) antibiotics usage in the 30 days preceding sample collection, and (E) patient's home geographic district in Uganda. Figure from (Monaco et al., 2016).

Adapted from (Monaco et al., 2016)

2.5 Many bacteria taxa are differentially abundant between cohort subsets grouped by peripheral CD4+ T cell counts

Using an adaptation of the differential expression computational tool DESeq2 modified to identify differentially abundant microbial taxa (Anders and Huber, 2010; Love et al., 2014; McHardy et al., 2013; McMurdie and Holmes, 2012), we were able to determine which bacterial operational taxonomic units (OTUs) were enriched or depleted in relation to subject peripheral CD4+ T cell count. These discriminant OTUs were then plotted by bacterial family association (**Figure 2.7A**). OTUs belonging to 13 bacterial families (*Enterobacteriaceae*, *Pasteurellaceae*, *Streptococcaceae*, *Enterococcaceae*, *Lactobacillaceae*, *Planococcaceae*, *Actinomycetalaceae*, *Carnobacteriaceae*, *Micrococcaceae*, *Gemellaceae*, *Comamonadaceae,* *Leuconostocaceae*, and *Leptotrichiaceae)* were enriched in HIV CD4<200 subjects in comparison to both HIV-uninfected subjects (**Figure 2.7A**) and HIV CD4>200 subjects (**Figure 2.7B**). OTUs assigned to the families *Succinovibrionaceae*, *Spirochaetaceae*, *S24-7*, and *Methanobacteriaceae* were depleted in HIV CD4<200 subjects in comparison to both HIV-uninfected subjects and HIV CD4>200 subjects. Other differentially abundant families included *Ruminococcaceae*, with fewer OTUs enriched than depleted in HIV CD4<200 subjects in comparison to the other two groups (56 OTUs depleted and 33 OTUs enriched in HIV CD4<200 versus HIV CD4>200; 50 OTUs depleted and 12 OTUs enriched in HIV CD4<200 subjects versus HIV-uninfected subjects) and *Bacteroidaceae*, with more OTUs enriched than depleted in HIV CD4<200 subjects in comparison to the other two groups (6 OTUs depleted and 31 OTUs enriched in HIV CD4<200 versus HIV CD4>200; 1 OTU depleted and 67 OTUs enriched in HIV CD4<200 versus HIV-uninfected subjects).

Adapted from (Monaco et al., 2016)

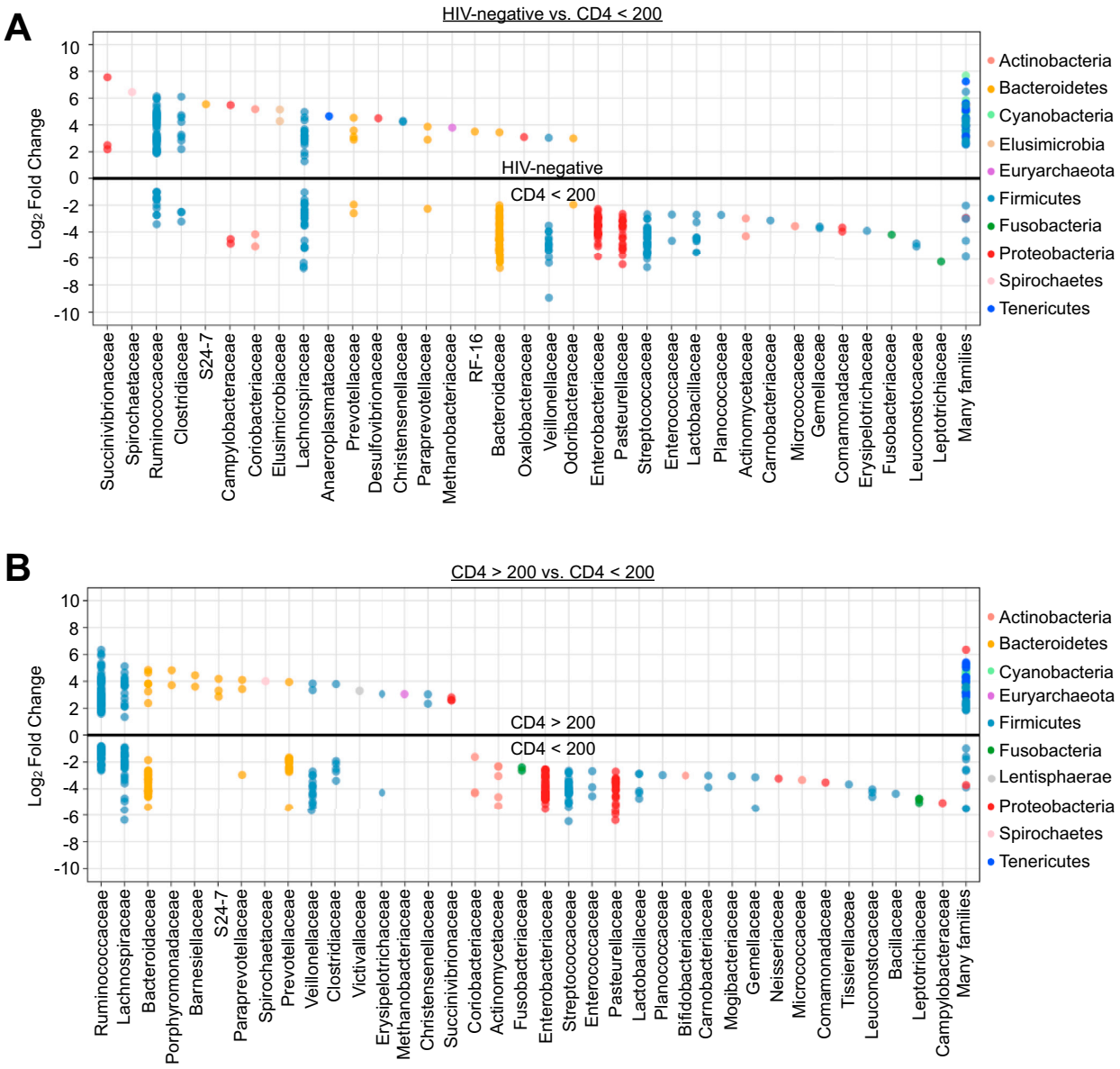

Figure 2.7. Differentially abundant bacterial taxa. DESeq2 was used to identify differentially abundant bacterial OTUs. Each dot represents one OTU assigned to the indicated bacterial family (column label). The position of the dot on the y-axis indicates the fold change of the enrichment (positive) or depletion (negative) of that OTU in relation to the subject cohort indicated in the top half of the panel. Bacterial OTUs that are significantly differentially abundant between (A) HIV-uninfected and HIV-infected subjects with peripheral CD4+ T cell count less than 200 as well as (B) HIV-infected subjects with peripheral CD4+ T cell count greater than 200 versus those with peripheral CD4+ T cell count less than 200 were graphed by $\log_2$ fold-change (y-axis) and grouped by family association (x-axis). Coloring indicates bacterial phyla to which the OTUs belong. Figure from (Monaco et al., 2016).

Adapted from (Monaco et al., 2016)

In order to determine the possible sub-family level taxonomic assignments within the differentially abundant *Ruminococcaceae* and *Bacteroidaceae* OTUs that might predict enrichment or depletion in relation to HIV status and peripheral CD4+ T cell count, we used the computational method oligotyping (Eren et al., 2011) to further refine the taxonomic assignment of *Ruminococcaceae* and *Bacteroidaceae* sequences that had not been originally assigned at a species level with QIIME. Briefly, oligotyping determines what combination of sequences contains most of the information present in all of the sequences associated with differentially abundant OTUs. From this subset of the larger pool of sequences, the most contributory more specific taxa assignments can be determined. Within the family *Ruminococcaceae,* those sequences most representative of OTUs depleted in HIV CD4<200 subjects compared to HIV CD4>200 subjects were assigned to the species *Ruminococcus bromii* and *Ruminococcus callidus* and those most representative of OTUs depleted in HIV CD4<200 subjects compared to HIV-uninfected subjects were assigned to the species *Ruminococcus bromii*. Species could not be confidently assigned to those OTUs enriched in HIV CD4<200 subjects or to differentially abundant OTUs in the family *Bacteroidaceae*.

2.6 Many bacterial taxa are independently differentially abundant between cohort subsets grouped by peripheral CD4+ T cell counts

Six metadata variables (month stool collected, co-trimoxazole use, age, MiSeq run number, antibiotic use, and BMI) were either significantly or near significantly different between patient subsets as determined by HIV status and peripheral CD4+ T cell count, so we performed multivariate modeling using the computational tool Multivariate Association with Linear Models (MaAsLin) (Morgan et al., 2012) to determine which

Adapted from (Monaco et al., 2016)

bacterial OTUs would be associated with HIV status and peripheral CD4+ T cell count independent of these potential confounders.

Since co-trimoxazole use was a correlated covariate with HIV status, it was excluded from the model. Fourteen bacterial OTUs were independently associated with peripheral CD4+ T cell count (**Table 2.2**) including OTUs assigned to *Ruminococcaceae* and *Enterobacteriaceae* families. Using the oligotyping technique (as previously described), we assigned the *Enterobacteriaceae* sequences to genus *Shigella* or a related *Escherichia* species but were unable to further assign the *Ruminococcaceae*. These data provide evidence for gut microbiome differences in a sub-Saharan African population with advanced HIV disease independent of other differentiating metadata factors.

Adapted from (Monaco et al., 2016)

Table 2.2: OTUs significantly associated with CD4 T cell number using a multivariate model including CD4 T cell group, age, month stool collected, sequencing run, BMI, and other antibiotic use. Table from (Monaco et al., 2016).

Phyla	Class	Order	Family	Genus	OTU ID	Raw *p* value	FDR *p* value
Firmicutes	Clostridia	Clostridiales	Clostridiaceae	NA	575768	9.52E-05	0.021
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Oscillospira	187267	0.000124	0.026
Bacteroidetes	Bacteroidia	Bacteroidales	NA	NA	343635	0.000202	0.035
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	782953	1.99E-05	0.007
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	111294	2.18E-05	0.007
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	292289	2.27E-05	0.007
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	110656	2.43E-05	0.007
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	435477	8.57E-05	0.020
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Cronobacter dublinensis	667570	0.00016	0.030
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	581782	0.000235	0.039
Firmicutes	Clostridia	Clostridiales	NA	NA	448176	0.000295	0.048
Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus parainfluenzae	4347099	0.000308	0.048
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	NA	199182	0.00033	0.048
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	888300	0.000331	0.048

Adapted from (Monaco et al., 2016)

2.7 Many bacterial taxa within prior families of interest are associated with plasma sCD14

As increased levels of circulating sCD14 have been linked to increased mortality in HIV (Sandler et al., 2011) and the protein may represent a proxy for bacterial translocation from the gut (Marchetti et al., 2013), we correlated bacterial OTUs with circulating sCD14 levels. We identified 144 bacterial OTUs significantly associated with sCD14, including 23 assigned to the family *Bacteriodaceae* and 23 assigned to the family *Ruminococcaceae*. Of those, OTU 199182 assigned to the family *Ruminococcaceae* and OTU 575768 assigned to the family *Clostridiaceae* were also associated with peripheral CD4+ T cell count independently of other metadata variables from the multivariate analysis above (**Table 2.2**). These associations suggests a possible relationship between these taxa and host inflammation.

2.8 Co-trimoxazole has a small effect on the gut microbiota in this cohort

The majority of HIV-infected subjects in our study were receiving co-trimoxazole (trimethoprim-sulfamethoxazole) prophylaxis (**Table 2.1**) per the World Health Organization and The Joint United Nations Programme on HIV and AIDS (WHO/UNAIDS) guidelines for HIV-infected individuals in resource-limited settings (World Health Organization, 2006). While there is extensive literature describing the effect of antibiotic use on the enteric microbiome (Blaser, 2016, 2018; Bokulich et al., 2016; Candon et al., 2015; Cox et al., 2014; Livanos et al., 2016; Nobel et al., 2015; Ruiz et al., 2017; Satokari et al., 2014; Schubert et al., 2015; Wlodarska et al., 2011), there has been only a small body of work examining the effect of co-trimoxazole specifically (de Bonnecaze et al., 2018; Hara et al., 2012; Kofteridis et al., 2004; Mavromanolakis et al., 1997; Oldenburg et al., 2018; Stamey et al., 1977). To

Adapted from (Monaco et al., 2016)

investigate the possible effect of long-term usage of co-trimoxazole on the gut microbiota of the subjects within this Ugandan cohort, we employed the previously described computational tool DESeq2 to identify differentially abundant bacterial OTUs with regard to co-trimoxazole use. As only four total HIV-infected subjects in our cohort were not receiving co-trimoxazole, we performed differential abundance testing on only HIV-uninfected controls. Within this patient subset, we found only two bacterial OTUs that exhibited significant fold-change with co-trimoxazole use. In comparison, we were able to detect 429 differentially abundant OTUs when comparing HIV CD4<200 subjects to those with peripheral CD4+ T cell count greater than 200. In addition, we did not find a difference in phylogenetic diversity of the gut microbiota between subjects receiving co-trimoxazole (n = 68) when compared to those that were not (n = 39, $p = 0.378$). In total, these data suggest that long-term co-trimoxazole treatment has a relatively small effect on the gut microbiota when compared to other co-occurring perturbations and that, within our cohort, extended use of co-trimoxazole would only contribute minimally to the enteric microbiome alterations observed in subjects with immunodeficiency.

2.9 Discussion

This study represents the first examination of HIV-associated gut microbiome differences conducted in sub-Saharan Africa, which is the region of the world that bears the greatest share of the global HIV burden (UNAIDS, 2018). The enteric microbial community changes we observed in the subjects with advanced HIV/AIDS mirrored previous findings in SIV-infected macaques (Handley et al., 2016; Handley et al., 2012), with decreased richness and phylogenetic diversity. Similarly, our observation that bacterial OTUs assigned to the family *Enterobacteriaceae* were associated with

Adapted from (Monaco et al., 2016)

decreased peripheral CD4+ T cell count is congruent with the finding that OTUs within *Enterobacteriaceae* were associated with a lower percentage of circulating CD4+ T cells in rhesus macaques (Handley et al., 2016). In our study, the oligotyping technique identified these organisms as genus *Shigella* or a closely related *Escherichia* species. This suggests a possible functional significance, as *Enterobacteriaceae* family members have been associated with inflammation (Lupp et al., 2007) and many members of the genus *Shigella* are enteric pathogens (Hale and Keusch, 1996; Yang et al., 2005). These microbial taxa could contribute to the gastrointestinal disease and chronic immune activation observed in HIV patients.

In contrast to the previously-described putative inflammatory activity of the OTUs within the family *Enterobacteriaceae*, increased levels of the genus *Ruminococcus* were associated with greater numbers of peripheral CD4+ T cells and decreased plasma sCD14 levels, meaning that these bacteria could be functioning in a protective role. These findings are concordant with several previous observations, including that *Ruminococcus* members are depleted in the rectal mucosa of HIV-infected, untreated subjects (McHardy et al., 2013) and in symptomatic SIV-infected macaques (Handley et al., 2016). *Ruminococcus* has been associated with both protective and disruptive roles in the enteric microbiome, and this effect is highly species dependent (Crost et al., 2013; David et al., 2014; Hsiao et al., 2014; Ze et al., 2012). Using oligotyping, we identified two species, *R. callidus* and *R. bromii*, depleted in HIV CD4<200 subjects. Both of these species have been shown to increase after therapeutic fecal microbial transplant treatment for *Clostridium difficile* infection (Satokari et al., 2014) and to be decreased in subjects with the gastrointestinal disorders Crohn's disease and ulcerative

Adapted from (Monaco et al., 2016)

colitis (Kang et al., 2010; Rajilic-Stojanovic et al., 2013), suggesting a protective role. Consistent with a possible protective role, the strong association of *R. bromii* with healthy microbial communities has led to its proposed use as a probiotic (Scott et al., 2015).

Prior studies of the enteric bacterial microbiome in HIV-infected individuals have reported disparate findings with regard to observed differences in the bacterial community. While there is substantial overall disagreement, some patterns have emerged, including enrichment of bacteria in the genus *Prevotella* and depletion of bacteria in the genus *Bacteroides* in chronic HIV infection (Dillon et al., 2014; Dinh et al., 2015; Lozupone et al., 2013; Lozupone et al., 2014; McHardy et al., 2013; Mutlu et al., 2014; Vazquez-Castellanos et al., 2014; Vujkovic-Cvijin et al., 2013). While we found significant differences in abundance of bacterial OTUs assigned to the genus *Prevotella* and *Bacteroides* when our cohort was subset by peripheral CD4+ T cell count, the directionality of these enrichments and depletions were not completely in concordance with those that have been reported in the literature. Of note, many characteristics varied between studies, including exclusion criteria, duration of ART treatment, sampling method and site, sample size, sequencing technique, and statistical methods, which could contribute to the discrepancies observed between this study and others. Most importantly, prior studies were performed on subjects in high-income countries with relatively well-controlled HIV (none with advanced HIV disease) while our cohort was from rural Uganda and included patients with AIDS. Differences in these host features are associated with some of the greatest variation measured in human gut microbiomes. Low-income, rural, and agrarian populations differ greatly from their

Adapted from (Monaco et al., 2016)

counterparts and are often characterized by high abundances of bacteria in the genus *Prevotella* and low abundances of bacteria in the genus *Bacteroides* (De Filippo et al., 2010; Smits et al., 2017). This is consistent with the average *Prevotella* abundance of 28% we observed in our HIV-uninfected subjects from a rural Ugandan population. It is possible that because the baseline gut microbiota of individuals in these populations is already *Prevotella*-rich and *Bacteroides*-poor, they would not experience *Prevotella*-associated changes observed with HIV infection in subjects from high-income countries. Together these results suggest that therapies targeting the enteric bacterial microbiome that are successful in HICs may not have the same efficacy in LICs. If this supposition is true, then it further increases the value and urgency of conducting comparative microbiome studies that include cohorts from regions of interest such as sub-Saharan Africa.

2.10 Limitations of this study

There were several limitations of this study. We did not have data about our subjects' co-morbidities so were unable to assess AIDS-defining infections and cancers. Dietary information was also unavailable, which could mean a potential confounder was unexamined, as diet can have a profound effect on the enteric bacterial and viral microbiome (David et al., 2014; Minot et al., 2011; Wu et al., 2011). Further, HIV-uninfected subjects were recruited to the study from a population of individuals that presented to the HIV clinic for care, so they may not represent healthy community controls. Due to obstacles related to specimen collection and preparation, we were unable to assess all components of the microbiome, including RNA viruses and intestinal parasites. As a result of these limitations, while we were able to probe the

Adapted from (Monaco et al., 2016)

DNA virome in the larger study (not covered in this dissertation) (Monaco et al., 2016) we were unable to explore facets of the microbiome that might be particularly relevant in HIV infection, as expansion of the RNA virome was observed in SIV-infected macaques that progressed to AIDS (Handley et al., 2016; Handley et al., 2012).

Chapter 3: HIV-associated Gut Microbiome Differences and Immune Activation are Dependent on Host Context

3.1 Investigating HIV, the enteric microbiota, and clinical correlates across the span of geographic variation

Our first characterization of HIV-associated enteric microbiome differences in sub-Saharan Africa provided a preliminary indication that the different microbial milieu within individuals in this region could influence the eventual HIV-associated enteric microbiome differences (Monaco et al., 2016). We observed the increased abundance of organisms in the family Prevotellaceae in HIV-uninfected subjects that was consistent with previous descriptions of gut microbial communities in individuals in LICs (De Filippo et al., 2010; Gomez et al., 2016; Gorvitovskaia et al., 2016; Lozupone et al., 2013; Martinez et al., 2015; Obregon-Tito et al., 2015; Schnorr et al., 2014; Smits et al., 2017; Wu et al., 2011; Yatsunenko et al., 2012). Because of the significant differences in baseline gut microbial communities in individuals in different geographic locations, we questioned the ability to translate studies conducted in HICs to the LICs that bore the greatest burden of HIV (UNAIDS, 2018). In light of our data, studies that supported a model proposing that HIV-infected individuals have increased *Prevotella* abundance and that this is associated with an elevated inflammatory state would seem to be ill suited to describe HIV-associated gut microbiome differences in a population that exhibits high *Prevotella* abundance at healthy baseline. Our first study examined HIV-associated microbiome differences at only one geographic location, so we were unable to determine how broadly these findings would be applicable across populations in sub-Saharan Africa. For the subsequent study, our aim was to assess cohorts within multiple locations in sub-Saharan Africa as well as in an HIC to allow us to more completely describe possible variations in HIV-associated microbiome differences between HICs and LICs. By way of comparisons within the three sites, we could

determine the amount of variation between the locations within Africa and compare it to the differences between the African cohorts and the North American cohort. This comparison of intra- and inter-continental microbial differences would allow us to assess to what extent conclusions from studies on HIV and the gut microbiome in HIC populations might be relevant to LIC populations. In order to conduct the study at these multiple locations, we again partnered with colleagues in the MGH Global Health Division to collect samples from the rural region surrounding the city of Mbarara in Southwestern Uganda. In addition, we worked with research clinicians at Harvard School of Public Health to analyze samples collected at Princess Marina Hospital in Gaborone, Botswana. Gaborone is the capital of and largest urban area in Botswana and Princess Marina is a large referral and central teaching hospital in Botswana with multiple international contacts. To establish an HIC cohort, we used the pre-existing resources associated with a Boston, MA-based cohort at the Ragon Institute of MGH, MIT, and Harvard.

In addition to the ability to make both intra- and inter-continental comparisons, the implementation of a new comparative study would allow several improvements over our prior work. Rather than recruiting our HIV-uninfected control subjects from an HIV testing clinic as we had done in (Monaco et al., 2016), we would be able to use robustly selected community-based HIV-uninfected controls (Muiru et al., 2018). Our fecal samples had been previously collected in RNAlater, which has been shown to interfere with the extraction of nucleic acid from samples, especially those protocols that use phenol-chloroform (Angel, 2012; Popova et al., 2010; Renshaw et al., 2015). With the initiation of a new study, we would be able to modify our protocols to preserve fecal

samples by immediate freezing and storage at -80˚C, which is a preferred method of sample preservation in the microbiome field (Blekhman et al., 2016). Lastly, further time for recruitment as well as multiple recruitment locations would allow us to increase the sample size and consequently the power of our study.

Here we examine the effect of HIV on the enteric bacterial microbiome in three distinct geographical locations spanning two continents as well as both HICs and LICs, representing regions that are particularly relevant to the current HIV epidemic. We demonstrate that there are unique HIV-associated gut microbial differences at each location, both at the whole-community and individual taxa levels. These microbial differences also coincide with differences in the relationship between microbial taxa and host inflammatory markers. We found that the sexual behavior of individuals in HICs is an interacting factor that can determine the directionality and magnitude of HIV-associated microbiome differences. Taken together, these findings suggest that HIV infection could alter anti-inflammatory microbial functions and allow the emergence of new functions. However all of these effects are highly context-dependent on the host, their pre-existing gut microbial community, and their geographic microbial milieu. This model has significant implications for efforts that intend to translate the results of HIV-microbiome studies between different geographical locations.

3.2 Cohort characteristics

We have access to large, well-characterized patient cohorts with clinical metadata based in Uganda (Hunt et al., 2011; Monaco et al., 2016; Muyanja et al., 2015; Okello et al., 2015; Siedner et al., 2016a; Siedner et al., 2016b; Siedner et al., 2013) and Botswana (Okatch et al., 2016), as well as a Boston, MA, USA-based cohort operated

with the resources of the Ragon Institute of MGH, MIT, and Harvard. In our Ugandan cohort, HIV-infected subjects have been stably on ART for at least 5 years and HIV-uninfected subjects are location-matched controls. The three cohort locations in our study represent an urban location in a HIC (Boston, MA, USA), an urban location in LIC (Gaborone, Botswana), and a rural location in a LIC (Mbarara District, Uganda) as measured by income, housing conditions, and demographics (**Table 3.1**). Gaborone is the capital of Botswana and a large urban center, while the subjects recruited from the Mbarara district resided in rural areas outside of the district capital. As a result of this geographical diversity, our sites span the range of many demographics that are proposed to underlie geography-associated enteric microbiome (**Table 3.1**), such as income or urbanization (Ayeni et al., 2018; Winglee et al., 2017).

Table 3.1. Patient characteristics and cardiovascular disease metadata for Boston, Botswana, and Uganda cohorts

For comparing continuous variables, Mann-Whitney and Kruskal-Wallis tests were used. For comparing categorical variables, Chi-Square and Fisher's exact tests were used.NA, not available or not applicable; NS, not significant; IQR, interquartile range; ART, antiretroviral therapy; MSM, men who have sex with men; CVD, cardiovascular disease; CIMT, carotid intima-media thickness; NRTI, nucleoside reverse transcriptase inhibitors; NNRTI, non-nucleoside reverse transcriptase inhibitors.

[g]Includes individuals with HIV both on ART treatment and not on ART treatment (i.e., elite controllers).

Table 3.1 (Continued)

	Boston cohort (*n* = 296)				Botswana cohort (*n* = 195)				Uganda cohort (*n* = *205*)		
Geographical Characteristics											
Per capita income (USD/month)	$5,490.83[a]				$850.12[b]				$159.47[c]		
Population density (/km^2)	4,793[d]				1,100[e]				265[f]		
	HIV infected virally suppressed[g] (*n* = 108)	**HIV infected not virally supressed** (*n* = 58)	**HIV uninfected** (*n* = 130)	***p* value**	**HIV infected on ART** (*n* = 75)	**HIV infected not on ART** (*n* = 40)	**HIV uninfected** (*n* = 80)	***p* value**	**HIV infected on ART** (*n* = 105)	**HIV uninfected** (*n* = 100)	***p* value**
Patient Characteristics											
Demographics, median (IQR)											
Age (years)	52 (46-58)	52 (39-54)	48 (30-55)	0.0015	39 (36-41)	37 (34-42)	38 (34-42)	NS	49 (45-51)	50 (46-54)	0.052
Male, *n* (%)	84 (90%)	39 (87%)	80 (71%)	0.001	35 (47%)	18 (45%)	43 (54%)	NS	51 (49%)	50 (50%)	NS
Height (cm)	178 (173-183)	175 (168-178)	175 (170-180)	0.011	166 (160-174)	165 (160-171)	167 (162-173)	NS	162 (158-170)	162 (157-168)	NS
Weight (kg)	84 (77-91)	82 (73-98)	79 (67-89)	0.023	68 (58-76)	61 (55 - 71)	67 (56-78)	NS	61 (54-66)	57 (51-66)	NS
Body mass index	26 (24-29)	28 (26-32)	25 (23-28)	0.017	24 (21-28)	22 (20-26)	23 (21-28)	NS	22 (20-25)	22 (19-25)	NS
Sexual behavior/practices, *n* (%)											
Non-MSM	26 (24%)	10 (17%)	83 (64%)		NA	NA	NA		NA	NA	
MSM	64 (59%)	41 (71%)	31 (24%)	<0.00001	NA	NA	NA		NA	NA	
Unspecified	18 (17%)	7 (12%)	16 (12%)		NA	NA	NA		NA	NA	
Laboratory measures, median (IQR)											
CD4 T cell count (cells/*μ*l)	775 (538-963)	579 (462-792)	933 (619-1189)	0.00018	515 (393-695)	353 (155-540)	NA	NS	430 (334-546)	NA	
>500, *n* (%)	78 (80%)	34 (64%)	24 (89%)		38 (51%)	11 (28%)	NA		33 (31%)	NA	
200-500, *n* (%)	16 (16%)	18 (34%)	2 (7%)	<0.00001	34 (45%)	18 (45%)	NA	0.0005	65 (62%)	NA	
<200, *n* (%)	3 (4%)	1 (2%)	1 (4%)		3 (4%)	11 (28%)	NA		7 (7%)	NA	
CD4 nadir (cells/*μ*l)	439 (333-654)	402 (349-592)	913 (552-1145)	<0.00001	128 (73-180)	NA	NA		122 (80-175)	NA	
Current ART, *n* (%)	65 (60%)	NA	NA		75 (100%)	NA	NA		105 (100%)	NA	
NRTI	44 (68%)	NA	NA		40 (53%)	NA	NA		105 (100%)	NA	
NNRTI	17 (26%)	NA	NA		35 (47%)	NA	NA		95 (90%)	NA	
Protease inhibitor (excludes ritonavir)	13 (20%)	NA	NA		6 (8%)	NA	NA		10 (10%)	NA	
Ritonavir	11 (17%)	NA	NA		0 (%)	NA	NA		0 (%)	NA	
Integrase inhibitor	20 (31%)	NA	NA		0 (%)	NA	NA		0 (%)	NA	
Pharmokinetic enhancer	11 (17%)	NA	NA		0 (%)	NA	NA		0 (%)	NA	
CCR5 antagonist	1 (2%)	NA	NA		0 (%)	NA	NA		0 (%)	NA	
Unspecified regimen	20 (31%)	NA	NA		25 (33%)	NA	NA		0 (0%)	NA	
Years on ART, median (IQR)	9.6 (4.3-14.6)	NA	NA		9.8 (8.0-11.8)	NA	NA		7.0 (6.4-7.5)	NA	
Cardiovascular Disease Metadata											
CVD laboratory measures, median (IQR)											
Triglycerides (mg/dL)	NA	NA	NA		46 (31-70)	43 (31-58)	35 (27-50)	NS	114 (91-139)	117 (96-156)	NS
High density cholesterol (mg/dL)	NA	NA	NA		54 (43-66)	43 (35-54)	54 (46-66)	NS	44 (37-53)	45 (39-50)	NS
Low density cholesterol (mg/dL)	NA	NA	NA		108 (81-131)	85 (66-104)	85 (70-116)	NS	74 (56-93)	87 (69-105)	0.004
Glycosylated hemoglobin (HbA1c %)	NA	NA	NA		5.3 (5.1-5.6)	5.6 (5.4-6.0)	5.5 (5.2-5.6)	NS	5.3 (5.0-5.7)	5.6 (5.3-5.9)	0.0008
Tobacco/cigarette use, current, *n* (%)	NA	NA	NA		27 (36%)	10 (24%)	28 (35%)	NS	4 (4%)	18 (18%)	0.001
Duration (years), median (IQR)	NA	NA	NA		7 (5-10)	10 (5-12)	10 (5-20)	NS	18 (6-28)	22 (12-28)	NS
Blood pressure (mmHg), median (IQR)											
Right systolic	NA	NA	NA		128 (116-137)	128 (117-136)	130 (120-137)	NS	112 (101-120)	120 (112-132)	9.2×10^{-5}
Left systolic	NA	NA	NA		126 (118-138)	128 (119-136)	129 (120-138)	NS	110 (100-121)	121 (111-135)	1.4×10^{-5}
Mean CIMT (mm), median (IQR)	NA	NA	NA		NA	NA	NA		0.63 (.58-.71)	0.68 (.63-.74)	0.0027
Right lateral	NA	NA	NA		0.41 (.37-.44)	0.39 (.35-.44)	0.40 (.36-.43)	NS	NA	NA	
Right anterior	NA	NA	NA		0.41 (.38-.44)	0.37 (.35-.43)	0.39 (.37-.43)	NS	NA	NA	
Right posterior	NA	NA	NA		0.41 (.39-.45)	0.39 (.37-.44)	0.40 (.37-.44)	NS	NA	NA	
Left lateral	NA	NA	NA		0.41 (.37-.45)	0.39 (.37-.43)	0.39 (.37-.43)	NS	NA	NA	
Left anterior	NA	NA	NA		0.41 (.37-.43)	0.40 (.36-.43)	0.40 (.37-.43)	NS	NA	NA	
Left posterior	NA	NA	NA		0.42 (.39-.45)	0.43 (.40-.46)	0.41 (.39-.45)	NS	NA	NA	

3.3 HIV-uninfected Individuals in different geographical locations have distinct gut microbiota

When examining human gut microbiomes on a global scale, variation in location is associated with some of the greatest variation in gut microbial community membership and structure (**Figure 1.4**). As stated previously, the gut microbiomes of subjects in HICs frequently feature higher abundances of organisms in the genus *Bacteroides, Bifidobacterium, and Ruminococcus* while those in LICs have higher abundances of organisms in the genus *Prevotella*, *Succinivibrio*, and *Treponema* (De Filippo et al., 2010; Gomez et al., 2016; Obregon-Tito et al., 2015; Schnorr et al., 2014; Smits et al., 2017; Yatsunenko et al., 2012). We characterized the gut microbiota of the subjects in the three cohorts using high-throughput sequencing of the V4 region (nucleotides 515-806) of the 16S rRNA gene and calculated the dissimilarity between the microbial communities of each individual. As could be predicted from the large differences in urbanization and geography (**Figure 1.3** and **Table 3.1**), the microbiota of HIV-uninfected individuals are distinct between each geographical location. The principal coordinate axis that accounts for the greatest proportion of the variation between individuals in these populations (PCo1) is characterized by a shift from microbial communities with a high proportion of Prevotellaceae in more rural locations to more heterogeneous communities in urban cohorts that feature a greater representation of members of the families Lachnospiraceae, Ruminococcaceae, and Bifidobacteriaceae (**Figure 3.1**). These differences between the microbial communities of subjects in each cohort are significant as determined by a PERMANOVA test applied to the inter-sample Jensen-Shannon divergence.

With regard to the complexity (alpha diversity) within each of the bacterial communities, those in Uganda have a significantly lower richness than those in either Botswana or America; there are no significant differences in Shannon diversity (accounting for both richness and evenness) between the three locations; and the Faith's phylogenetic diversity (accounting for richness, evenness, and phylogenetic relatedness) is significantly greater in Botswana than in either of the other two cohorts (**Figure 3.2**).

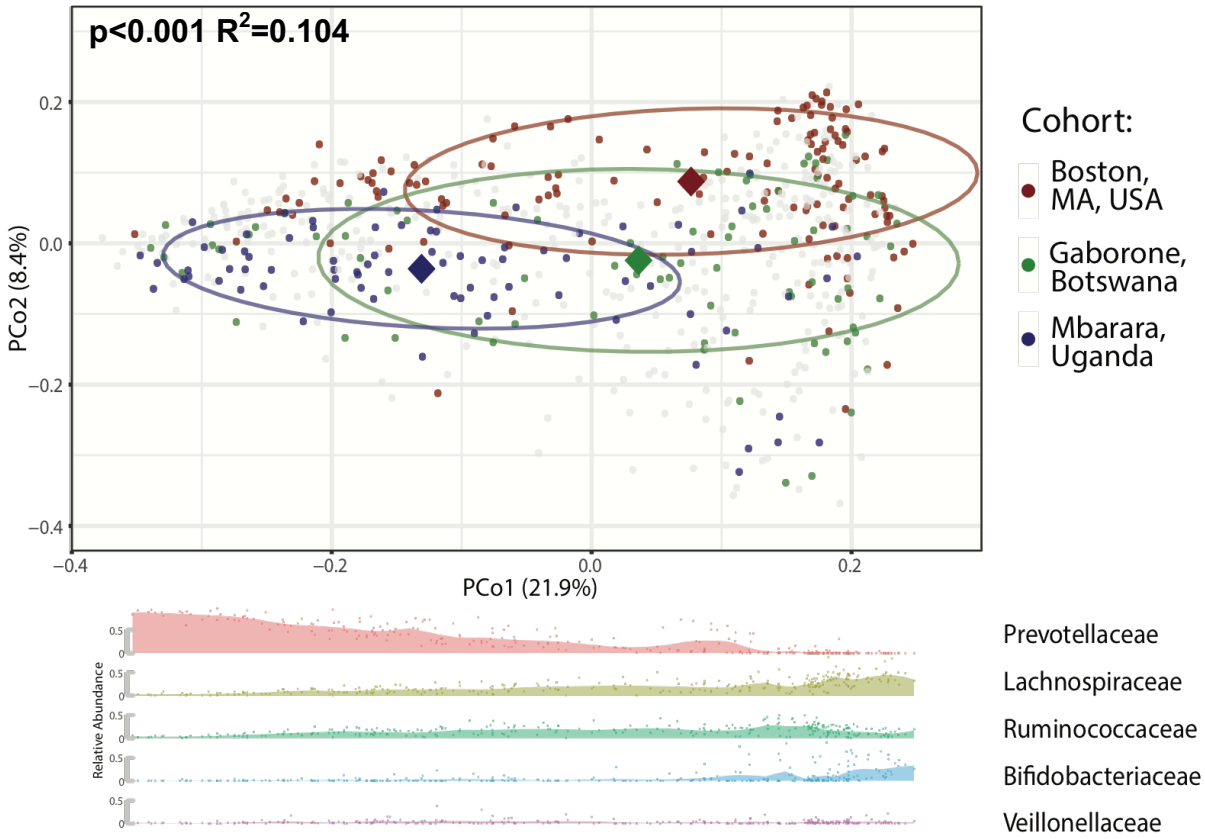

Figure 3.1. Gut microbiota of HIV-uninfected individuals are significantly different across geography. Inter-sample Jensen-Shannon divergence was calculated for all samples in the study and projected onto a 2-dimensional coordinate plane using PCoA. Only HIV-uninfected individuals are assigned colors; gray points are HIV-infected subjects. Ellipses are drawn at 68% confidence interval and centroids are noted with a diamond (◆) of the appropriate color. P value and R^2 are calculated by PERMANOVA of distance matrix by geographical location. Line graphs below the x-axis are of relative abundances of the 5 most abundant bacterial families based on projecting points from the main panel onto the x-axis.

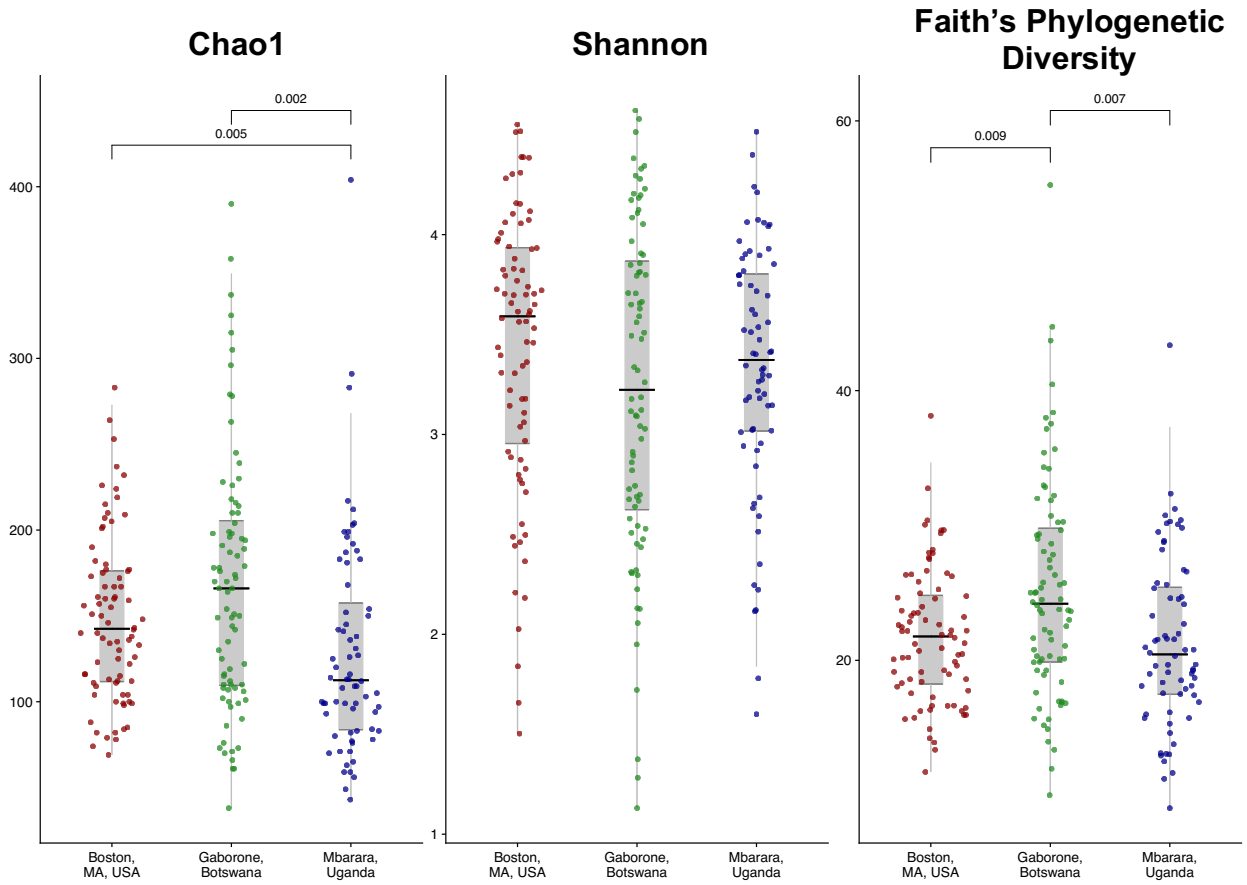

Figure 3.2. Alpha diversity of HIV-uninfected subjects at three geographical locations. Alpha diversity metrics of Chao1 (richness), Shannon Diversity (richness and evenness), and Faith's Phylogenetic Diversity (richness, evenness, and phylogenetic relatedness) were calculated on the 16S rRNA gene count matrix. Each point represents one sample and behind the points is a boxplot indicating middle two quartiles. The black horizontal bar represents the median value. P-value calculated by Wilcoxon rank sum test.

We can also determine the specific amplicon sequence variants (ASV) that are differentially abundant in one geographic location compared to another. ASVs each represent a unique 16S rRNA gene amplicon and are equivalent to organizational taxonomic units (OTUs) that are clustered at 100% identity (Callahan et al., 2017). The determination of exact ASV sequences is made possible by correcting for the error introduced by high-throughput sequencing (Callahan et al., 2016). We used the analysis tool DEseq2 (Love et al., 2014), which is well-suited to the zero-inflated nature of microbial count data (McMurdie and Holmes, 2014), to determine which microbial ASVs were differentially abundant between subjects in each geographic cohort. Because similar microbial communities will not have many significant differences in microbial abundance, the number of differentially abundant ASVs between two subject populations gives a measure of the dissimilarity between these two populations. In addition, knowing the identity of the enriched and depleted ASVs can allow the prediction of possible functional differences between the two microbial communities. Among HIV-uninfected subjects, we found 200 ASVs differentially abundant between Boston and Mbarara, 105 ASVs differentially abundant between Gaborone and Mbarara, and 100 ASVs differentially abundant between Boston and Gaborone (**Figure 3.3**), suggesting that the Ugandan and American subjects' microbial communities are more dissimilar to each other than either is to the Botswanan communities.

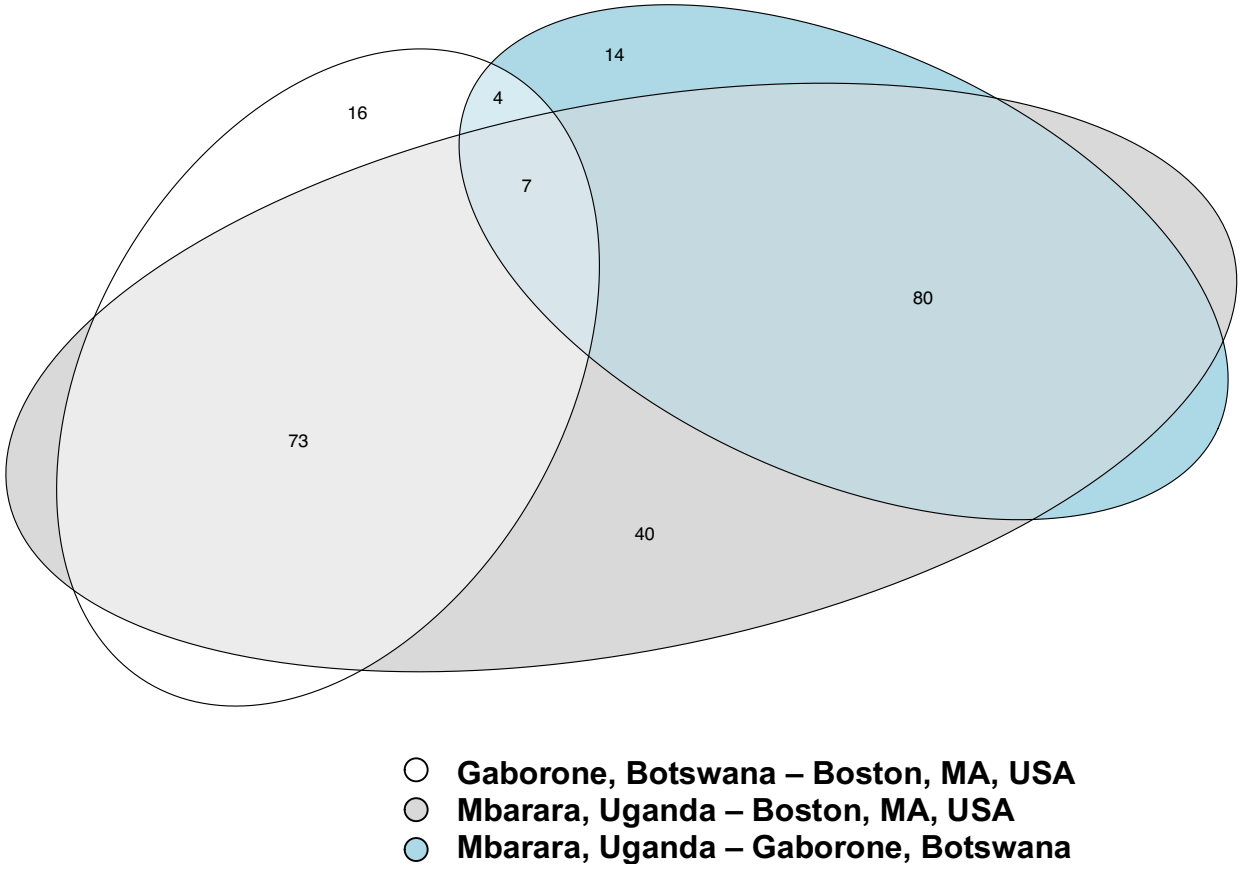

Figure 3.3. Quantity of differentially abundant ASVs between HIV-uninfected subjects in each geographical cohort. Differentially abundant ASVs between each geographic cohort were calculated pairwise using DESeq2. The total differentially abundant ASVs between two cohorts are contained within the appropriately colored ellipse. ASVs that appear within multiple pairwise comparisons with concordant directionality (e.g. an ASV that is enriched in Mbarara, Uganda relative to both Gaborone, Botswana and Boston, MA, USA) are counted in the appropriate ellipses overlap.

By identifying the overlap between the pairwise differences between cohorts, we can determine which ASVs uniquely identify each geographic location relative to the other two. Boston is characterized by 73 differentially abundant ASVs in comparison to the other two cohorts and Mbarara is characterized by 80 differentially abundant ASVs. Gaborone is characterized by 4 differentially abundant ASVs, far fewer than the other two cohorts, implying overlap with each of the other two cohorts (**Figure 3.3**). This result is due to the great discordance between those ASVs that are differentially abundant between Botswana and America and those that are differentially abundant between Botswana and Uganda.

We can appreciate the differences among HIV-uninfected subjects at each geographic location at a sample-by-sample level by visualizing those ASVs that are differentially abundant between the three cohorts (**Figure 3.4**). There are many ASVs that are present in all three groups but are differentially abundant that appear in the center rows of the heat map. Conversely, the ASVs at the top and bottom rows of the heat map include many ASVs that are completely absent from the American or Ugandan cohort. In contrast, there are few ASVs absent from the Botswanan cohort, and almost none absent from that cohort exclusively, suggesting that the Botswanan microbiomes could be considered a set of intermediate communities, with ASV membership that overlaps substantially with both the other two cohorts. Overall, the baseline gut microbial communities of HIV-uninfected individuals at the three geographic locations mimic those described in the literature, with greater abundances of *Prevotella* in the most rural sites within the LICs.

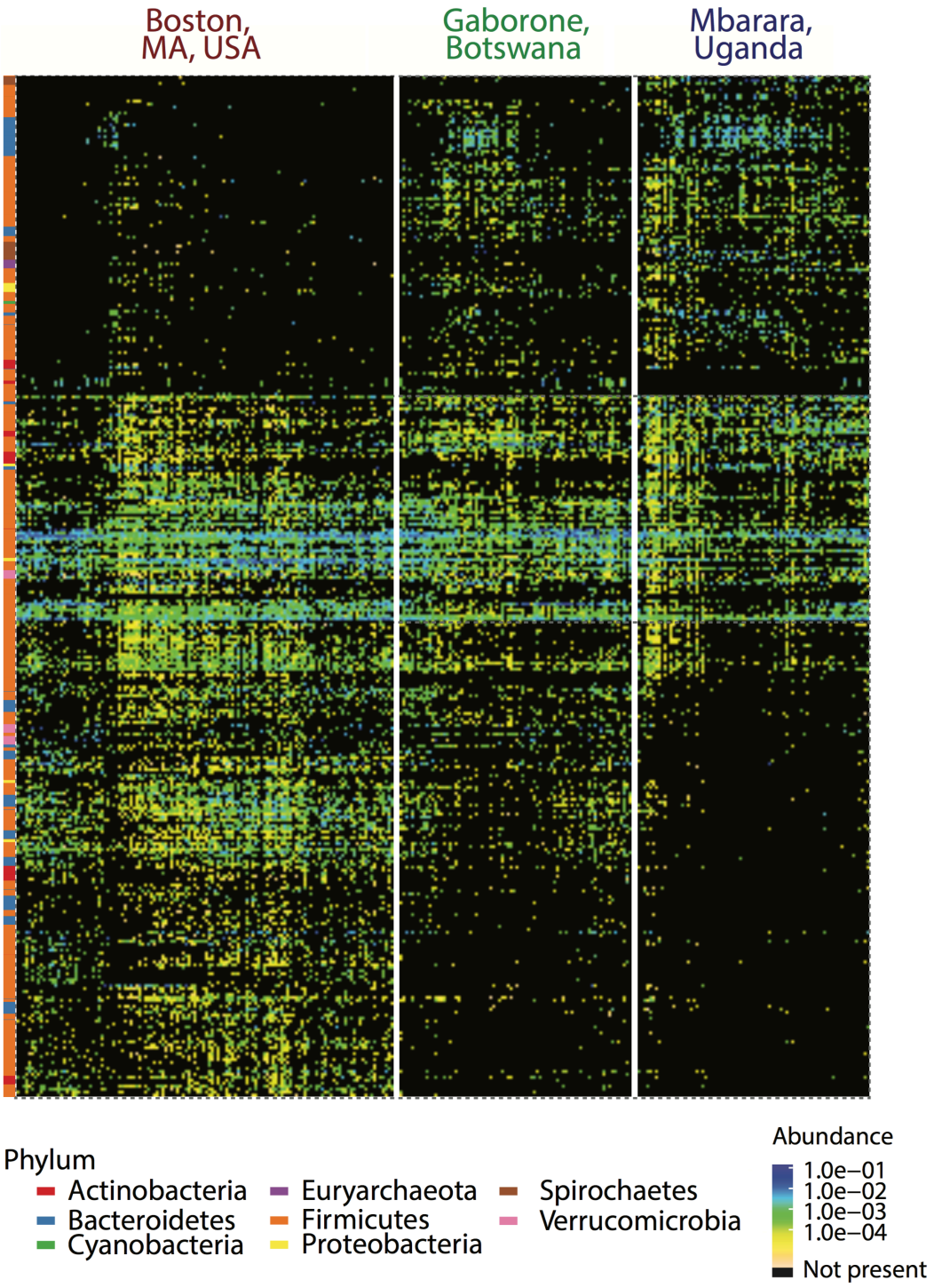

Figure 3.4. Heat map of differentially abundant ASVs by geography in HIV-uninfected subjects. Differentially abundant ASVs between each geographic cohort were calculated pairwise using DESeq2. Subjects are arrayed on the x-axis and divided into geographical cohort. ASVs are arrayed on the y-axis and their Phylum is indicated by coloring on the left side of the figure.

3.4 Circulating markers of inflammation are elevated in HIV-infected subjects

As systemic inflammation is hypothesized to be a contributing factor in many chronic complications and causes of additional mortality in HIV, we sought to determine whether the HIV-infected individuals in our cohorts exhibited increased levels of inflammation. We measured the previously-mentioned established circulating markers of inflammation (sCD14 and IL-6), macrophage activation (sCD163), and intestinal epithelial damage (FABP2) (Brenchley and Douek, 2012; Brenchley et al., 2006; Federico et al., 2009; Haller et al., 2000; Jiang et al., 2009; Kim et al., 2016; Reimund et al., 1996; Rogler and Andus, 1998). We found that, compared to HIV-uninfected individuals, in HIV-infected individuals sCD14 levels were elevated in all three cohorts, FABP2 levels were elevated in Boston and Gaborone (**Figure 3.5**), and both IL-6 and sCD163 were elevated in only Boston (data not shown). These results show that our HIV-infected subjects were exhibiting the expected increase in systemic inflammation, and that this inflammation may have been related to the activation of macrophages by microbial products such as LPS, as sCD14 was the most consistently elevated plasma marker.

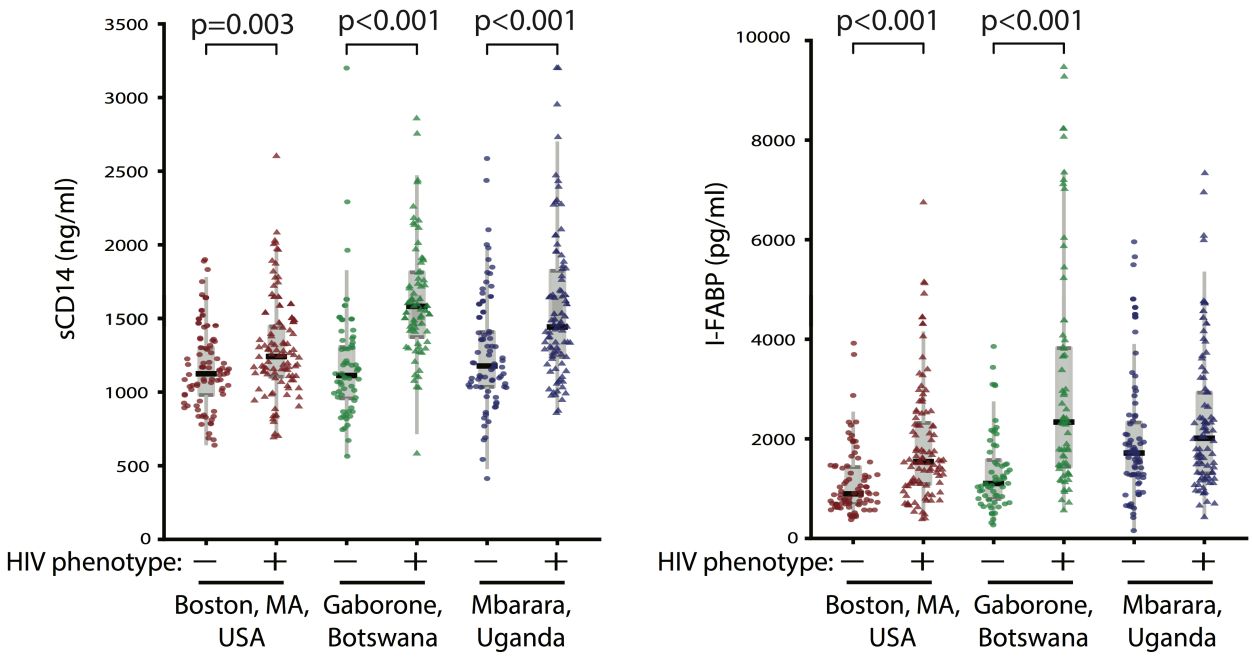

Figure 3.5. Many circulating inflammatory markers are elevated in HIV-infected subjects. Inflammatory markers were measured from patient plasma using immunological assays. The marker corresponding to each graph is displayed on the y-axis. Each point represents one sample and behind the points is a boxplot indicating middle two quartiles. The black horizontal bar represents the median value. P-value calculated by Wilcoxon rank sum test.

3.5 HIV-associated microbiota differ by geography and baseline microbial context

Having established and described the distinct baseline microbiota that are present in each of our cohorts, we can then evaluate the differences between the microbiota of HIV-infected and -uninfected subjects in this context. When plotting on the same coordinate system that was previously used to differentiate geographically-disparate communities (**Figure 3.1**), the gut microbial communities of HIV-infected and -uninfected subjects are significantly different within each cohort as calculated by the PERMANOVA test, which takes into account all principal coordinate axes (**Figure 3.6**). In the American and Ugandan populations, there are striking differences with regard to HIV infection status that are apparent in a principal coordinates projection and occur along the same PCo1 axis that distinguished geographically disparate HIV-uninfected subjects. However, the patterns of microbial changes that underlie these significant HIV-associated differences are unique to each cohort. In Boston, an urban setting within a HIC, HIV infection is associated with a more homogeneous overall community structure at the family level that has less abundance of members of the bacterial families Bifidobacteriaceae, Lachnospiraceae, and Ruminococcaceae and a greater abundance of members of the family Prevotellaceae. Conversely, in Mbarara, a rural setting in an LIC, HIV infection is associated with a more heterogeneous community with greater abundances of the bacterial families Bifidobacteriaceae and Lachnospiraceae and less abundance of Prevotellaceae. In Gaborone, which represents the middle of the urbanization and economic spectrum as an urban setting in an LIC, HIV-associated differences are subtler and not as apparent on the whole community level, though they are significantly different when calculated by PERMANOVA.

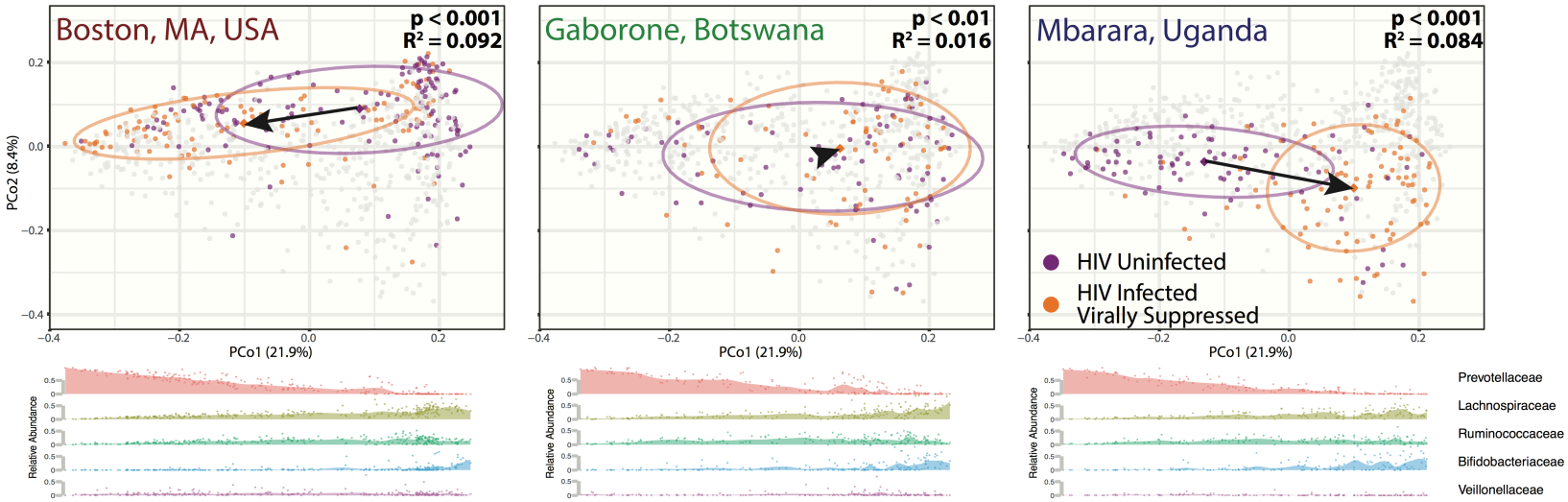

Figure 3.6. Gut microbiota are significantly different between HIV-uninfected and -infected subjects in each geographic cohort. Inter-sample Jensen-Shannon divergence was calculated for all samples in the study and projected onto a 2-dimensional coordinate plane using PCoA. In each panel, only subjects from the corresponding geographic cohort are assigned colors; gray points those subjects from the other two cohorts. Ellipses are drawn at 68% confidence interval and centroids are noted with a diamond (◆) of the appropriate color. An arrow indicates the positional change from the centroid of HIV-uninfected subjects to the centroid of HIV-infected subjects. P value and R^2 are calculated by PERMANOVA of distance matrix by geographical location. Line graphs below the x-axis are of relative abundances of the 5 most abundant bacterial families based on projecting points from the main panel onto the x-axis.

The family-level compositional differences are also apparent in fractional abundance graphs of the microbial communities of each subject (**Figure 3.7**). These differences are not strongly borne out in metrics of alpha diversity calculated on the gut microbial communities of HIV-infected and -uninfected subjects. When observing measures of richness and diversity, there are no significant HIV-associated pairwise changes in any of the three cohorts in either Chao1 (richness), Shannon diversity (richness and evenness), or Faith's phylogenetic diversity (richness, evenness, and phylogenetic relatedness) (**Figure 3.8**).

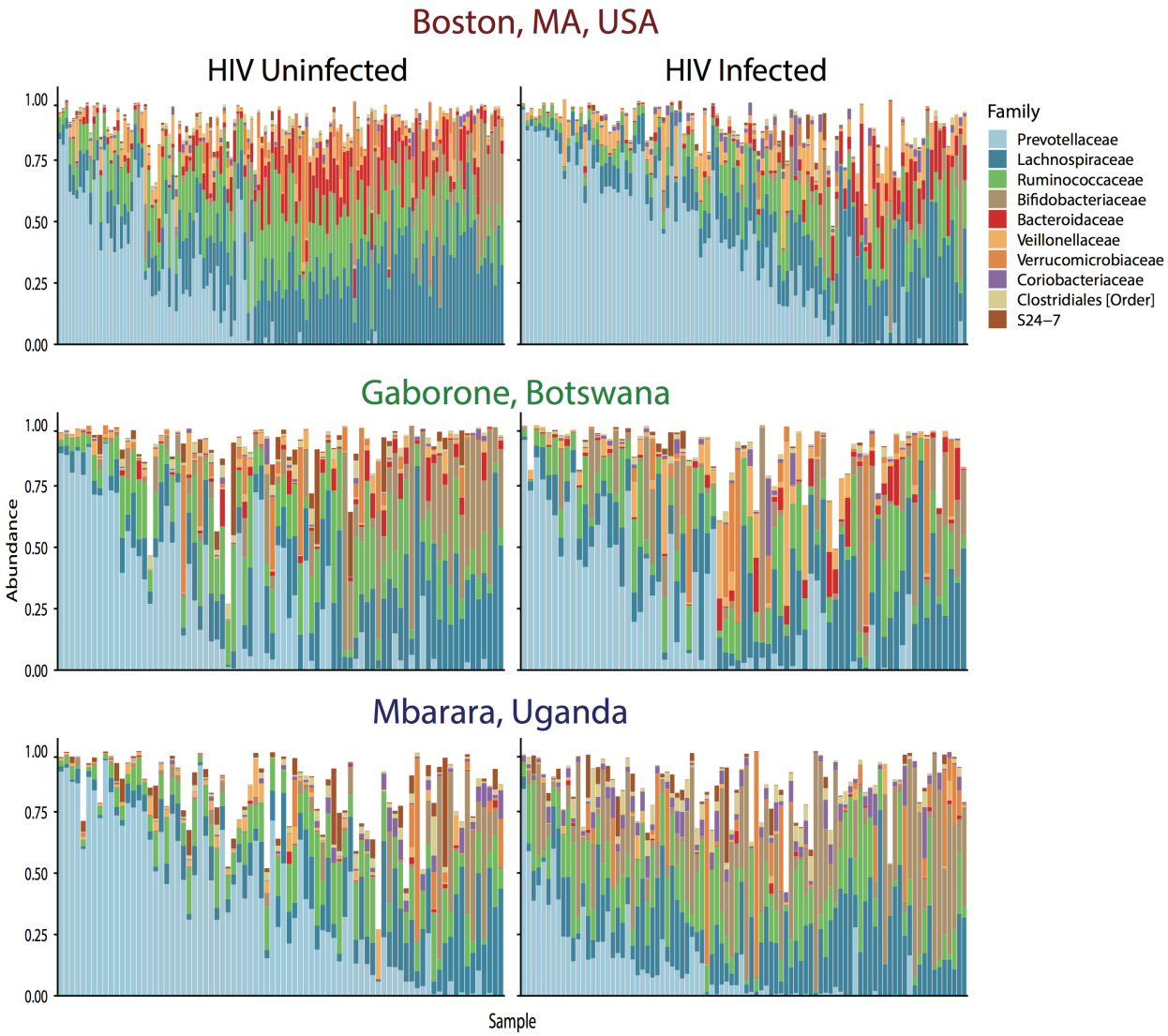

Figure 3.7. Fractional abundance composition of subjects' microbiomes at the family level. Subjects' samples are arrayed on the x-axis. Graphs in columns share the same HIV phenotype and graphs in rows share the same geographic cohort. Colors represent the 10 most abundant bacterial families. Reads that are not assigned to these families are left blank, meaning some bars will not reach the 1.00 fraction.

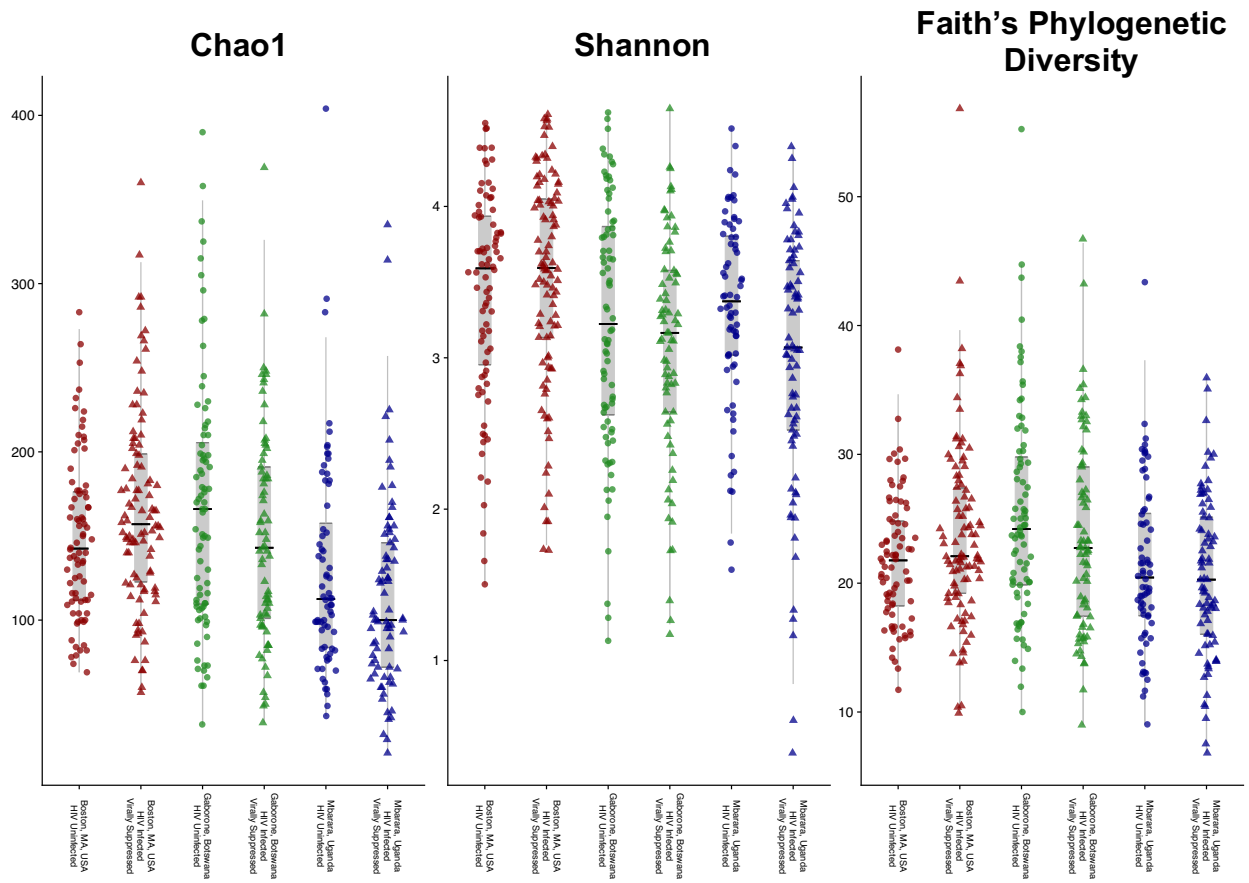

Figure 3.8. Alpha diversity of subjects separated by HIV phenotype at three geographical locations. Alpha diversity metrics of Chao1 (richness), Shannon Diversity (richness and evenness), and Faith's Phylogenetic Diversity (richness, evenness, and phylogenetic relatedness) were calculated on the 16S rRNA gene count matrix. Each point represents one sample and behind the points is a boxplot indicating middle two quartiles. The black horizontal bar represents the median value. P-values (none significant) calculated by Wilcoxon rank sum test.

As previously mentioned, the enteric microbiomes of men who have sex with men (MSM) have been found to be enriched in *Prevotella* at baseline, and this could possibly be confounding the HIV-associated microbial signal observed in HICs (Armstrong et al., 2018; Fulcher et al., 2018; Kehrmann et al., 2019; Kelley et al., 2017; Noguera-Julian et al., 2016; Nowak et al., 2019; Nowak et al., 2017; Pescatore et al., 2018). In order to work toward deconvoluting these two factors, we constructed our cohort in America, an HIC, to have representation from subjects across the combinatorial categories of sexual behavior and HIV phenotype. With these subjects, we were able to quantify the HIV-associated microbial differences in MSM and non-MSM individuals separately, essentially controlling for the influence of MSM. Plotting on the same principal coordinate axes of Jensen-Shannon divergence used previously shows that among non-MSM subjects, there were much weaker, but still significant, HIV-associated differences at the whole-community level ($p=0.021$ $R^2=0.017$), with both HIV-infected and HIV-uninfected subjects having microbial communities with low abundances of *Prevotella* (**Figure 3.9**). However, in MSM subjects, there were strong and significant HIV-associated microbial differences ($p<0.001$ $R^2=0.047$), with the HIV-uninfected phenotype characterized by intermediate levels of *Prevotella* and the HIV-infected phenotype characterized by high levels of *Prevotella*. Through the use of a large cohort with good subpopulation coverage, we were able to examine the relationship between MSM, HIV, and the microbiota in finer resolution than many previous studies. These findings suggest that MSM status and HIV phenotype are behaving as interacting factors, in that MSM behavior modifies the outcome of HIV-associated differences, but

is not a confounder entirely responsible for the previously-detected signal of HIV-associated microbiome differences.

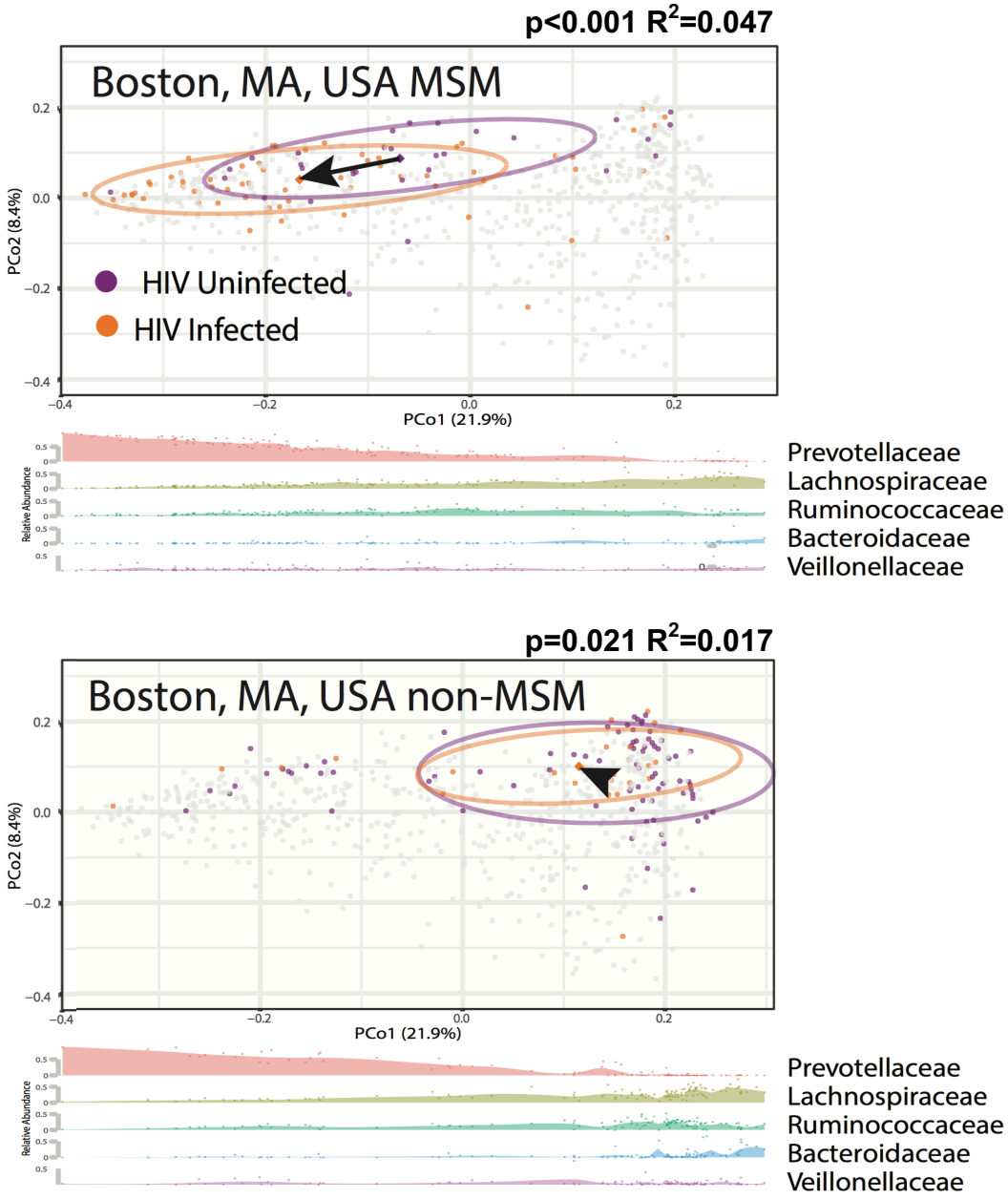

Figure 3.9. Gut microbiota are much more significantly different between HIV-uninfected and -infected subjects within the MSM subset than between those within the non-MSM subset in Boston, MA, USA. Inter-sample Jensen-Shannon divergence was calculated for all samples in the study and projected onto a 2-dimensional coordinate plane using PCoA. In each panel, only subjects from the corresponding geographic cohort are assigned colors; gray points are those subjects from the other two cohorts (Botswana and Uganda). Ellipses are drawn at 68% confidence interval and centroids are noted with a diamond (◆) of the appropriate color. An arrow indicates the positional change from the centroid of HIV-uninfected subjects to the centroid of HIV-infected subjects. P value and R^2 are calculated by PERMANOVA of distance matrix by geographical location. Line graphs below the x-axis are of relative abundances of the 5 most abundant bacterial families based on projecting points from the main panel onto the x-axis.

3.6 Diet does not strongly influence microbiota composition in Botswana

We had detailed diet information of macro- and micronutrient consumption for the subjects in the Botswanan cohort. As diet has been shown to have the potential to have a substantial effect on the microbiome (David et al., 2014; Wu et al., 2011), we aimed to determine whether the measured dietary constituents were associated with microbial community state. Characterizing this relationship would allow us to control for diet as a confounding factor if it did have a significant correlation with the microbiome.

The 38 dietary nutrient measurements were transformed by the addition of a pseudocount of one and then log transformed to moderate the left skew in the original data (**Figure 3.10**). Then, a principal components projection including normalization was performed to extract a mathematical combination of the nutrients that would encompass the greatest amount of variation within the diet data in only one measure. This produced a PC1 component that correlated well with many nutrients (**Figure 3.11**), contained a substantial amount of the variation within the original data (44.7%), and widely distributed each subjects' diet in order to facilitate better analysis (**Figure 3.12**). We took the position along this axis as a "diet score" for these individuals and used it to characterize differences and relationships related to diet. The diet score was not significantly different between the different HIV phenotype groups, suggesting that at a whole-diet level the diets of the members of each group were similar (**Figure 3.13**). In addition, including the diet score in a PERMANOVA of the Botswanan microbial communities did not show a relationship between those two factors ($p = 0.121$ and $R^2 = 0.0075$). By this analysis in this cohort, we did not find a strong correlation between diet and the enteric microbiome.

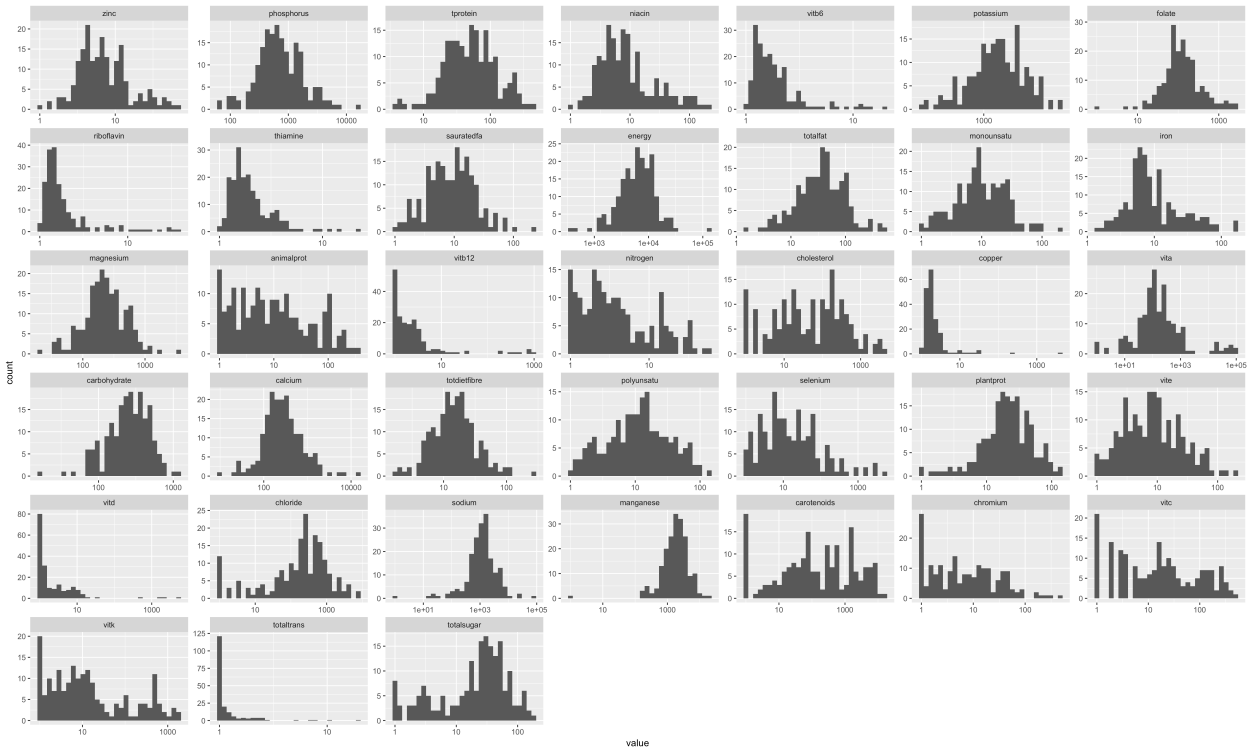

Figure 3.10. Histograms of Botswanan dietary nutrient distributions after pseudocount and log transformation. Original data on 38 nutrients from Botswanan subjects was transformed with the addition of a pseudocount of one and log transformation to increase the symmetry of the data in preparation for performing a principal components analysis.

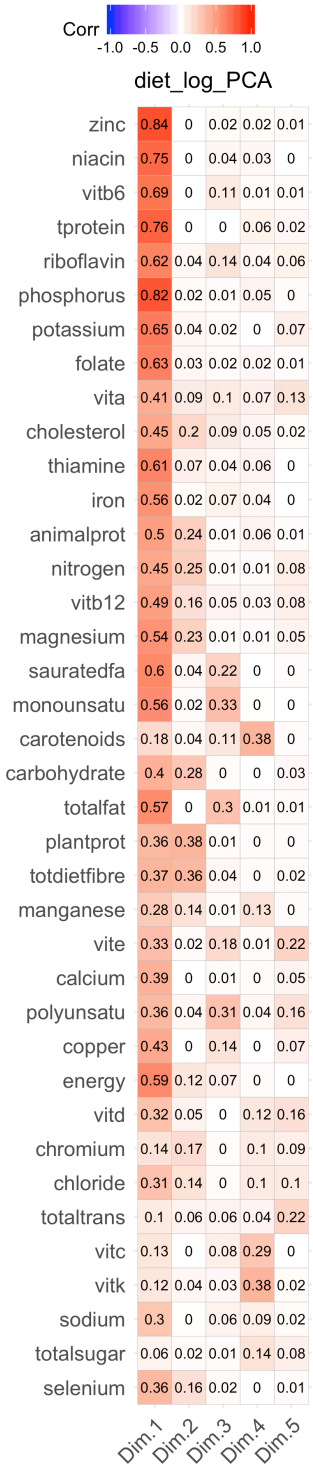

Figure 3.11. Principal component 1 correlates with many nutrients in the original Botswanan diet data. Spearman rank correlation between the first 5 dimensions of the principal component analysis. Dimension 1 (PC1) correlates highly with many of the most important and well-distributed nutrients, signifying that PC1 adequately describes the diet of each subject with only one continuous value.

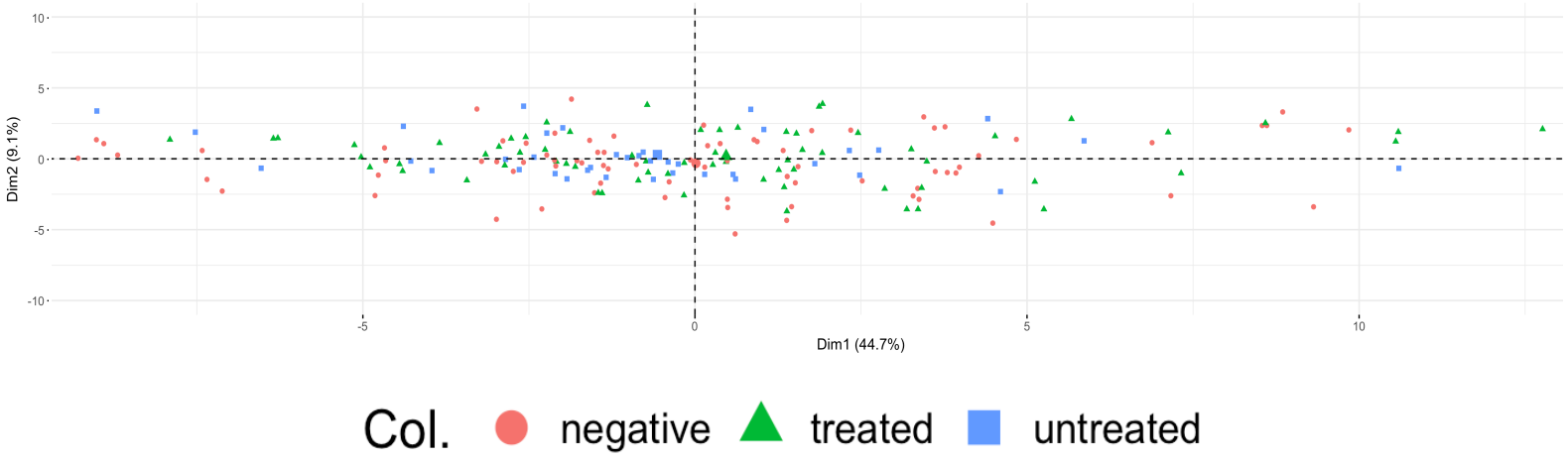

Figure 3.12. Principal components projection of nutrient intake of Botswanan subjects. Graph of the first two dimensions of the principal components analysis of transformed Botswanan nutrient values. Subjects are well distributed along the first axis (Dim.1). Subjects are colored by HIV status with HIV-uninfected (“negative”), HIV-infected and ART-treated (“treated”), and HIV-infected and ART-untreated (“untreated”) subjects shown.

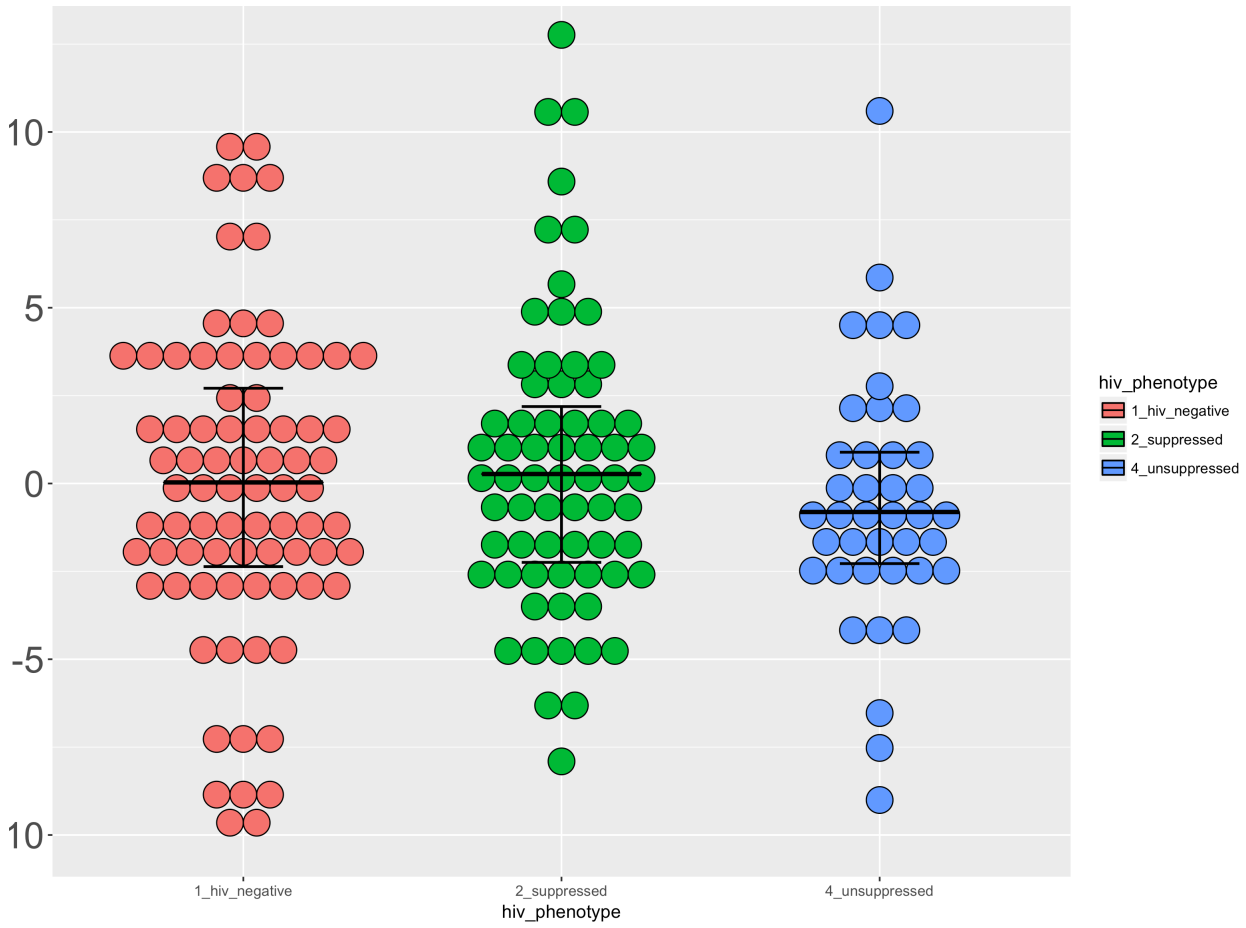

Figure 3.13. Derived Botswanan "diet score" is not significantly different between subjects grouped by HIV status. "Diet score" derived from principal component 1 in **Figure 3.12** plotted for each subject separated by HIV status. Shown are HIV-uninfected ("1_hiv_negative"), HIV-infected and ART-treated ("2_suppressed"), and HIV-infected and ART-untreated ("4_unsuppressed") subjects. Widest horizontal black bar indicates median and narrower horizontal black bars indicate upper and lower quartile. Diet score is not significantly different between the three groups by Kruskal-Wallis test ($p < 0.546$).

3.7 Cohorts are characterized by unique enriched and depleted taxa when measured at high taxonomic resolution

While whole-community distance metrics and family-level abundances of microorganisms give us a broad view of the significant differences between our subject groups according to geography and HIV phenotype, we can use ASVs, which are based on exact 16S rRNA gene V4 sequences, to describe more fine-scale taxonomic differences between our groups of interest. Using DESeq2 analysis of count data of 16S ASVs as previously described, we identified multiple ASVs that were enriched or depleted in association with HIV within each cohort (**Figures 3.14** and **3.15**). Out of the total 4495 ASVs found in all samples, the number of HIV-associated differentially abundant ASVs that were unique to each cohort was 100 in Boston (across 42 species assignments), 14 in Gaborone (across 13 species assignments), and 45 in Mbarara (across 31 species assignments). 35 ASVs in 16 species assignments were differentially abundant in more than one cohort (*i.e.* "shared" between cohorts). The greatest overlap of HIV-associated differentially abundant ASVs was between the Ugandan and American cohort, which shared 32 ASVs that were differentially abundant. However, most of these (30 of 32 ASVs) had discordant HIV-associations in Boston and Mbarara, *i.e*. those that were increased in HIV-infected individuals in Boston were decreased in HIV-infected individuals in Mbarara and vice-versa. This observation mirrors the whole-community findings that Ugandan and American microbial communities moved in opposite directions in principal coordinate space under an HIV-driven perturbation. The previous whole-community finding of a less strong association between microbial community and HIV phenotype in Botswana than the other cohorts was concordant with the observation that only 14 ASVs were differentially abundant with

regard to HIV in that cohort, in contrast to the 100 and 45 ASVs identified in Boston and Mbarara, respectively. ASVs assigned to genus *Prevotella*, and more specifically to the species *copri* within that genus, were some of the most commonly occurring ASVs (**Figure 3.14**). The American differentially abundant ASVs include 18 assigned to *Prevotella* of which 9 were assigned to *copri*; in Uganda 15 were assigned to *Prevotella* and 6 of these to *copri*; in Botswana 2 were assigned to *Prevotella* and 1 of these to *copri*; and among shared ASVs, 16 were assigned to *Prevotella* of which 8 were *copri* (**Figure 3.14**). In light of the previously described associations between *Prevotella* and HIV, MSM behavior, and geographical location, the frequent occurrence of members of this genus among HIV-associated differentially abundant ASVs in our cohorts led us to additionally investigate this genus.

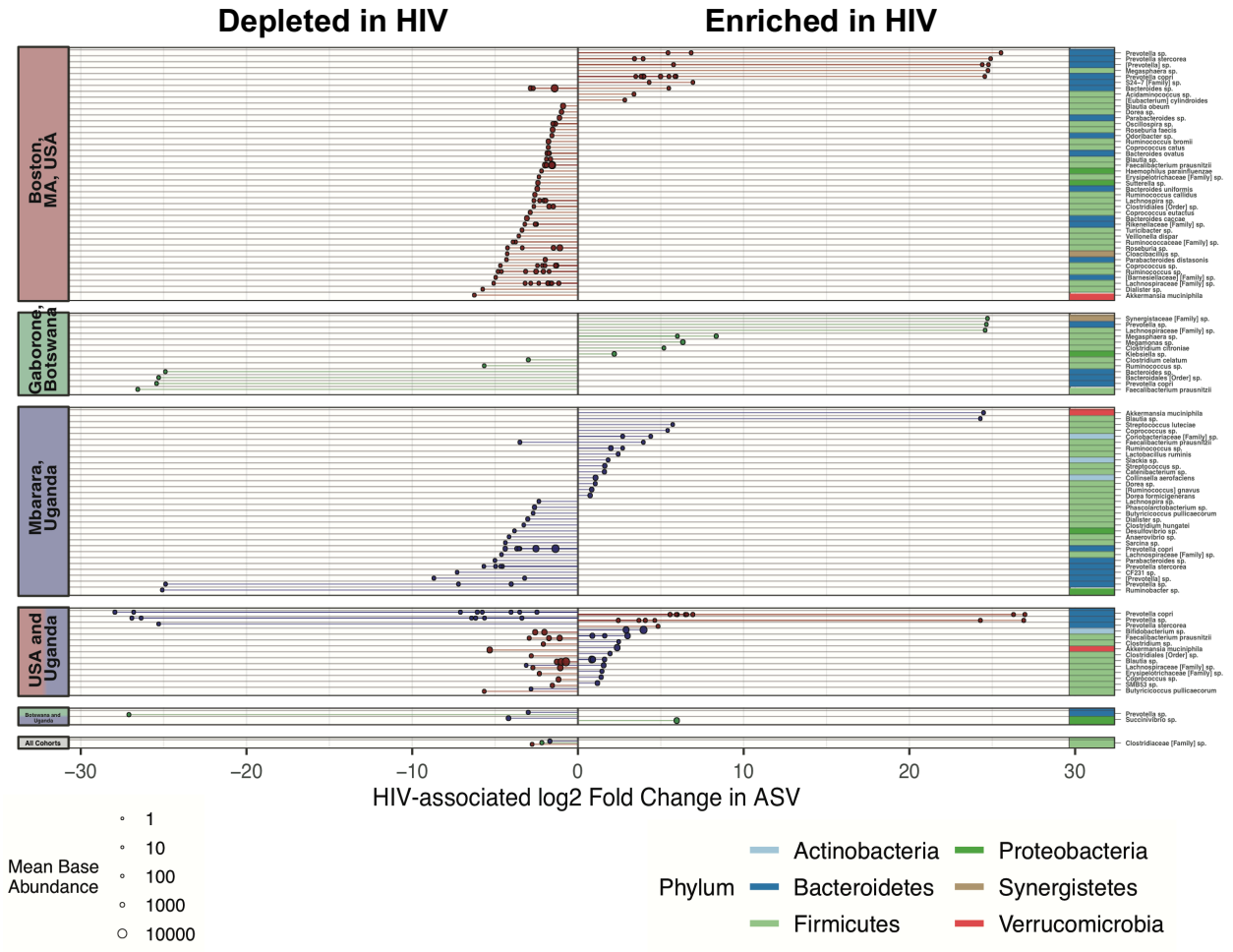

Figure 3.14. ASVs differentially abundant by HIV status grouped by geographic cohort occurrence. Differentially abundant ASVs between HIV-uninfected and -infected subjects within each cohort were calculated using DESeq2. Each dot represents one ASV. ASV dot color denotes the cohort it was differentially abundant in and ASV dot size represents normalized abundance. ASVs that are differentially abundant in only one cohort are displayed in one of the corresponding top three boxes; those that are differentially abundant in more than one cohort are displayed in one of the corresponding bottom three boxes. ASVs assigned to the same binomial name are collapsed and displayed as multiple dots on the same line. The phylum of the assigned binomial name is displayed on the right side of the figure alone with the assigned binomial name. ASVs present in multiple cohorts are represented with the appropriate number of dots, as fold change and abundance is different in each cohort.

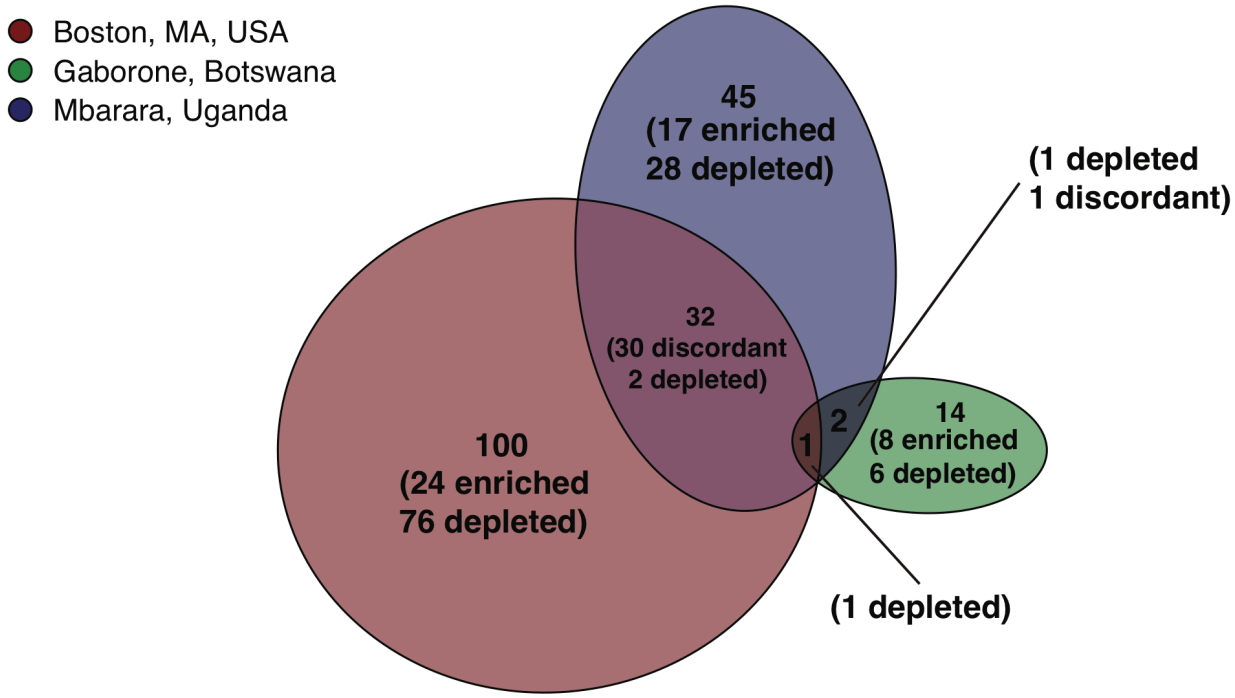

Figure 3.15. Euler diagram summarizing overlap in differentially abundant ASVs by HIV status for each cohort. Summary of differentially abundant ASVs displayed in **Figure 3.10**. Differentially abundant ASVs between HIV-uninfected and -infected subjects within each cohort were calculated using DESeq2 and the total ASV count is displayed within the appropriately colored ellipse. ASVs that are differentially abundant in more than one cohort are counted in the corresponding intersection area. ASVs that are more abundant in HIV-infected subjects are referred to as "enriched," while those that are more abundant in HIV-uninfected subjects are labeled "depleted." In intersections, ASVs are labeled "discordant" if the directionalities of the HIV-associated differential abundance do not match, *e.g.* if that ASV is enriched with regard to HIV infection in one cohort but depleted in another.

3.8 Microbiota-inflammation relationships are modified by geography

Beyond describing the HIV-associated enrichment and depletion of certain taxa, we aimed to further probe how these taxa associated with clinically relevant measures within our subjects. We chose to determine which ASVs were correlated with the previously mentioned markers of inflammation that predict mortality in HIV infection. We tested the association between the levels of these markers and the abundances of ASVs using DESeq2 to determine which ASVs were differentially abundant between the subjects in each cohort that were in different quartiles of measured levels of inflammatory markers.

For sCD14, the number of differentially abundant ASVs that were unique to each cohort was 9 in Boston (across 6 species assignments), 18 in Gaborone (across 12 species assignments), and 27 in Mbarara (across 16 species assignments) (**Figures 3.16** and **3.17**). There were 3 differentially abundant ASVs shared between the cohorts across 3 species assignments. The other three inflammatory markers FABP2, IL-6, and sCD163 were not associated with differential abundances of any ASVs. Similar to the finding from ASVs associated with HIV status, ASVs taxonomically assigned to the genus *Prevotella* were well represented among ASVs associated with differences in sCD14 (**Figure 3.16**). The American differentially abundant ASVs include 4 assigned to *Prevotella* of which none were assigned to *copri*; in Uganda 9 were assigned to *Prevotella* and 2 of these to *copri*; in Botswana 9 were assigned to *Prevotella* and 6 of these to *copri*; and among shared ASVs, 1 was assigned directly to *Prevotella copri*. In almost every case, *P. copri* has a negative association with sCD14 levels, suggesting that it could be acting in an anti-inflammatory role. Interestingly, there was no association between *P. copri* and sCD14 in the Boston cohort, meaning that in this

geographical location it does not appear to be filling any role similar to that proposed above. In addition, there were many ASVs that were unique to each cohort rather than shared among the cohorts, implying that the inflammatory associations and possible functions were geographically-dependent. Interestingly, the relationships between sCD14 and ASVs were dissimilar to those between CD4+ T-cell count and bacterial ASVs (**Figure 3.18**), possibly demonstrating that peripheral CD4+ T-cell count is not a good measure of microbial-associated immune activation in HIV, especially in chronically infected patients.

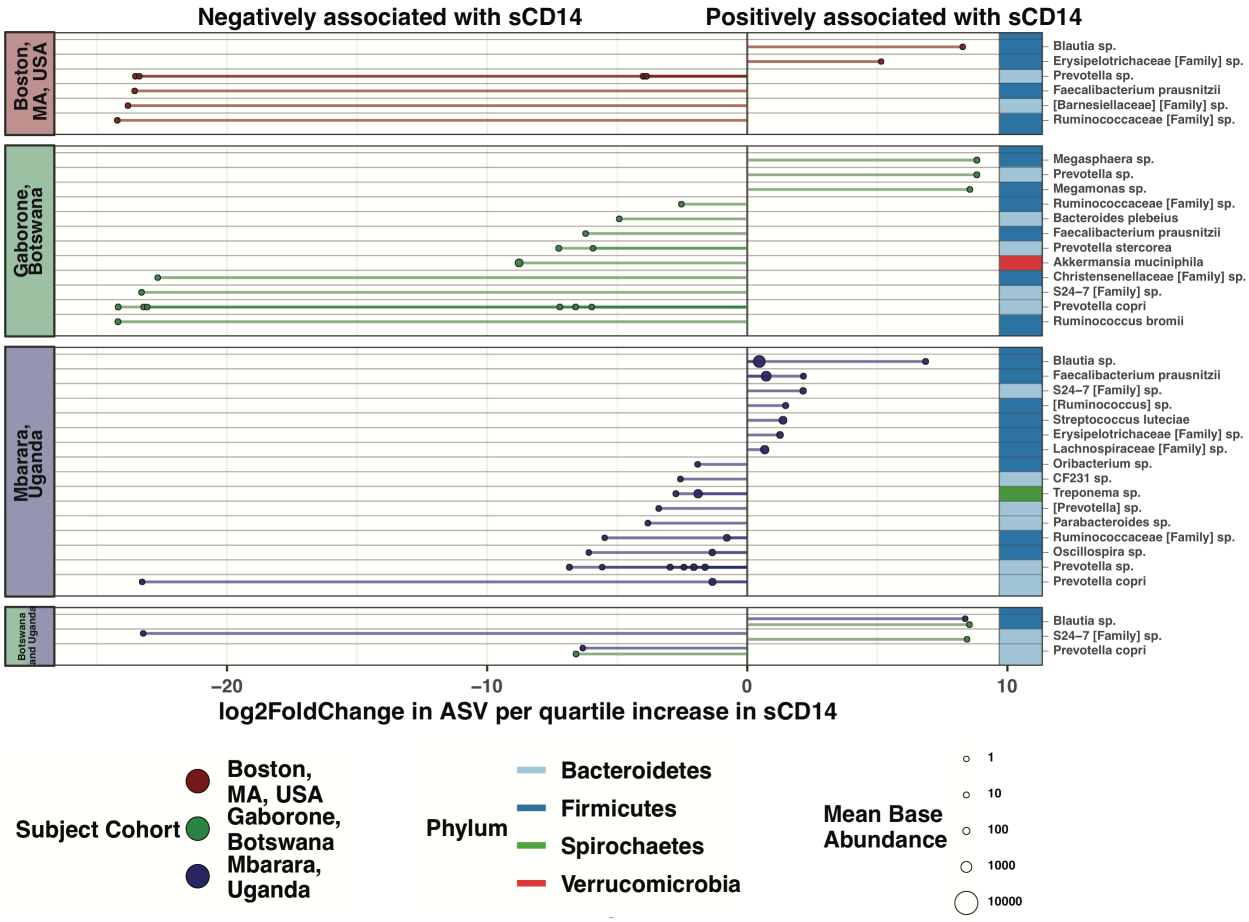

Figure 3.16. ASVs differentially abundant by sCD14 quartile grouped by geographic cohort occurrence. Differentially abundant ASVs between subjects in different sCD14 quartile within each cohort were calculated using DESeq2. Each dot represents one ASV. ASV dot color denotes the cohort it was differentially abundant in and ASV dot size represents normalized abundance. ASVs that are differentially abundant in only one cohort are displayed in one of the corresponding top three boxes; those that are differentially abundant in more than one cohort are displayed in one of the corresponding bottom three boxes. ASVs assigned to the same binomial name are collapsed and displayed as multiple dots on the same line. The phylum of the assigned binomial name is displayed on the right side of the figure alone with the assigned binomial name. ASVs present in multiple cohorts are represented with the appropriate number of dots, as fold change and abundance is different in each cohort.

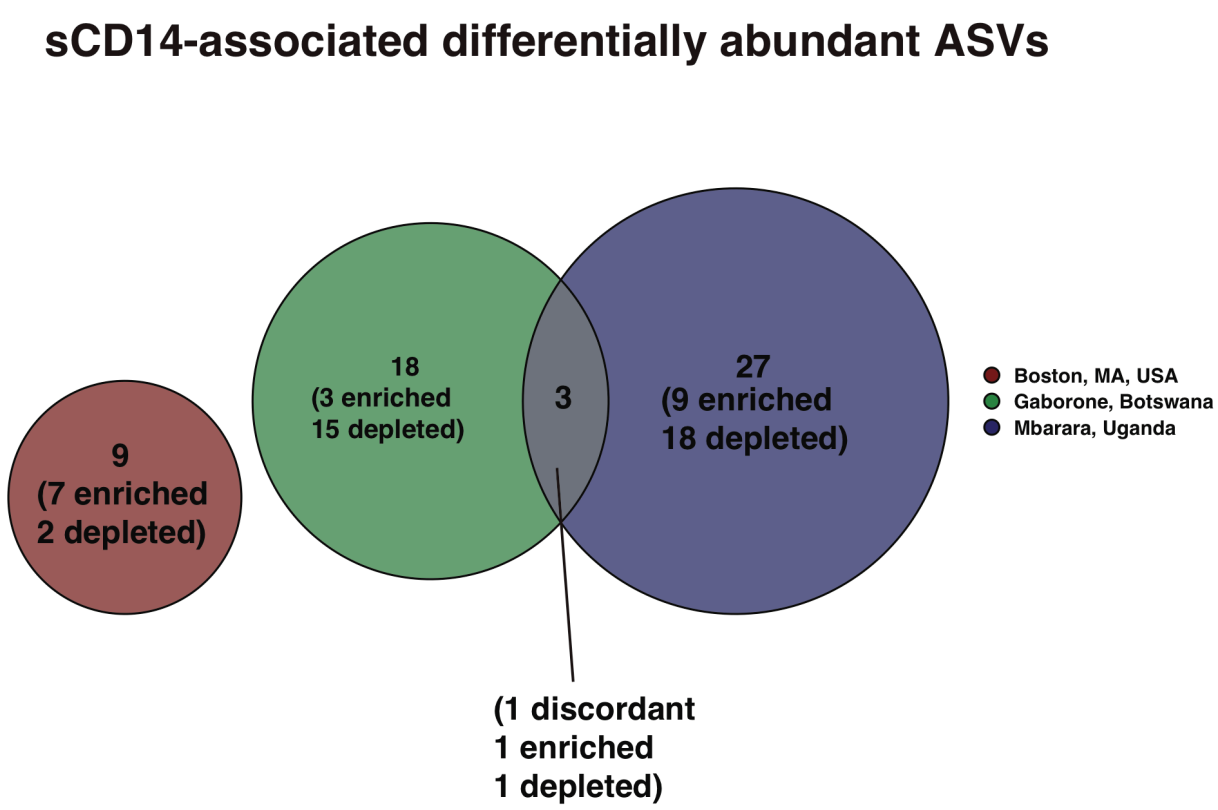

Figure 3.17. Euler diagram summarizing overlap in differentially abundant ASVs by sCD14 quartile for each cohort. Summary of differentially abundant ASVs displayed in **Figure 3.16**. Differentially abundant ASVs between subjects in different sCD14 quartile within each cohort were calculated using DESeq2 and the total ASV count is displayed within the appropriately colored ellipse. ASVs that are differentially abundant in more than one cohort are counted in the corresponding intersection area. ASVs that are more abundant in HIV-infected subjects are referred to as "enriched," while those that are more abundant in HIV-uninfected subjects are labeled "depleted." In intersections, ASVs are labeled "discordant" if the directionalities of the HIV-associated differential abundance do not match, *e.g.* if that ASV is enriched with regard to HIV infection in one cohort but depleted in another.

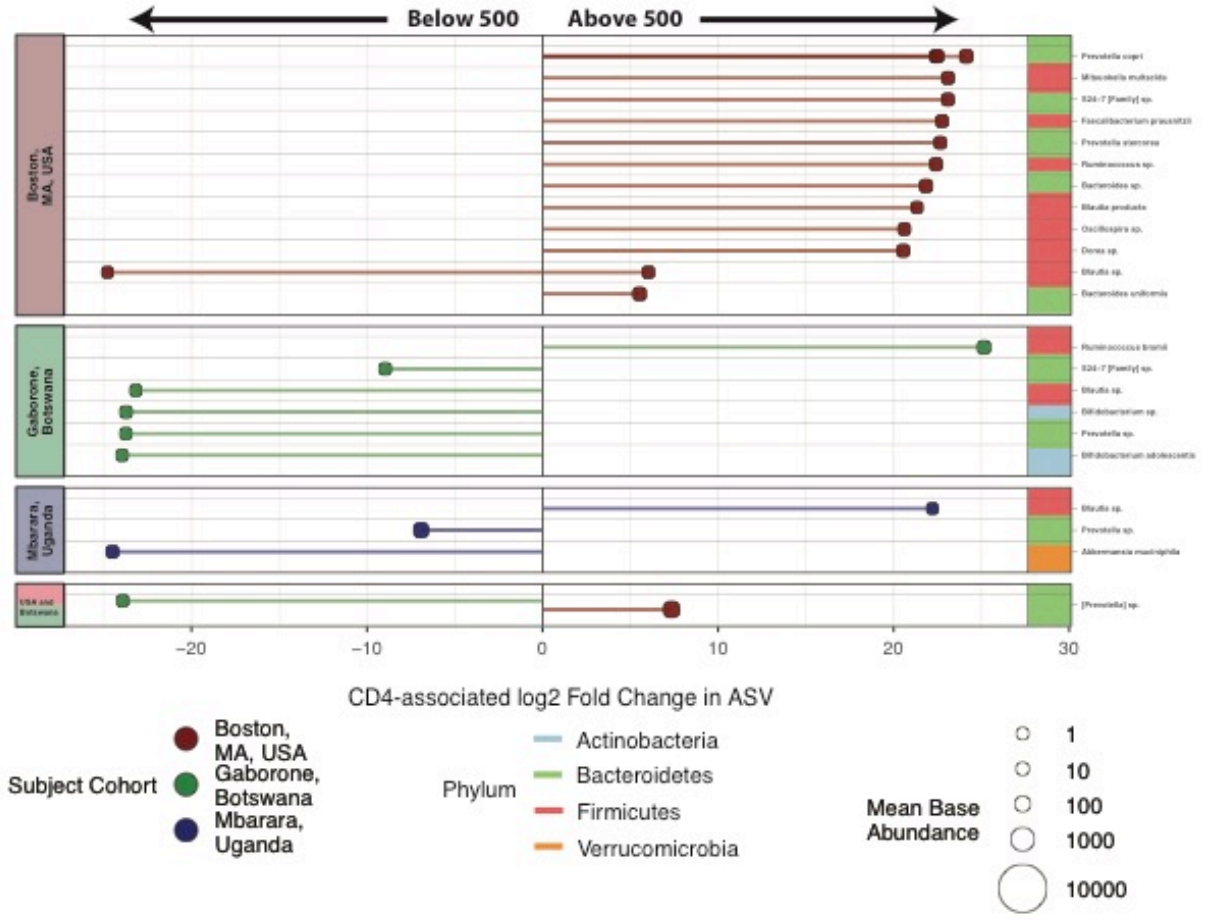

Figure 3.18. ASVs differentially abundant by peripheral CD4+ T-cell count above or below 500 cells/μl grouped by geographic cohort occurrence. Differentially abundant ASVs between HIV-infected subjects with peripheral CD4+ T-cell count above or below 350 cells/μl within each cohort were calculated using DESeq2. Each dot represents one ASV. ASV dot color denotes the cohort it was differentially abundant in and ASV dot size represents normalized abundance. ASVs that are differentially abundant in only one cohort are displayed in one of the corresponding top three boxes; those that are differentially abundant in more than one cohort are displayed in one of the corresponding bottom three boxes. ASVs assigned to the same binomial name are collapsed and displayed as multiple dots on the same line. The phylum of the assigned binomial name is displayed on the right side of the figure alone with the assigned binomial name. ASVs present in multiple cohorts are represented with the appropriate number of dots, as fold change and abundance is different in each cohort.

3.9 *Prevotella copri* from different geographic contexts possess different genomic functional capacities

Based on the patterns of associations with HIV infection and host inflammation we observed in our data, we decided to concentrate our exploration on bacteria within the genus Prevotella, specifically *Prevotella copri*. *Prevotella*, both in our study and in previous work, have been linked to host factors such as MSM behavior and geographical location that are particularly relevant in the context of HIV infection. As mentioned previously, members of the genus *Prevotella* have been found in many cases to be correlated with HIV infection and they have been proposed as potentially having a pro-inflammatory role in HIV-infected individuals, although there is a great deal of conflicting data regarding these conclusions (Duvallet et al., 2017; Lozupone et al., 2013; Lozupone et al., 2014; Noguera-Julian et al., 2016).

Also contributing to our decision to focus on *P. copri* were our findings that *P. copri* represented a substantial fraction of the *Prevotella* in the subjects in our studies (**Figure 3.19**), in concordance with prior work showing that it generally accounts for a substantial majority of *Prevotella* in the gut (Franke and Deppenmeier, 2018; Human Microbiome Project, 2012; Ley, 2016). In addition, *P. copri* has previously been connected with inflammatory conditions like rheumatoid arthritis (Scher et al., 2013) and osteolytic bone disease (Lukens et al., 2014), demonstrating that it may have inflammatory potential. In addition, only RefSeq contains only 22 sequenced *P. copri* genomes, so it is likely there is still meaningful uncharacterized diversity within the species. *P. copri*'s association with many dietary and metabolic interactions add to a diverse range of functions attributed to *P. copri*, and this diversity in function and occurrence has made it a good

candidate for examination at high taxonomic resolution (De Filippis et al., 2019; Metwaly and Haller, 2019).

To this end, we anaerobically isolated 115 single clones of *P. copri* from samples originating from HIC subjects in Boston, MA, USA and LIC subjects in Mbarara, Uganda. Our goal was to identify genomic variation that might account for observed functional differences as well as provide insight into the global diversity of *P. copri* by using isolates from two disparate geographic locations. After isolation and high-throughput sequencing, the *P. copri* genomes were assembled and the sequence of the 16S rRNA gene was determined for each isolate. We matched this sequence with the 16S rRNA gene ASVs derived from the subjects in our cohorts in order to establish which ASVs in the original 16S rRNA gene amplicon sequencing could have been contributed by our *P. copri* isolates. This would provide us with an approximation of the fraction of the enteric bacterial community represented by our isolates. We found that ASVs identical to those in our isolates accounted for a large proportion of the bacterial communities in our subjects (**Figure 3.19**), making it more probable that our isolate-derived conclusions can be extrapolated to the function of the total bacterial community.

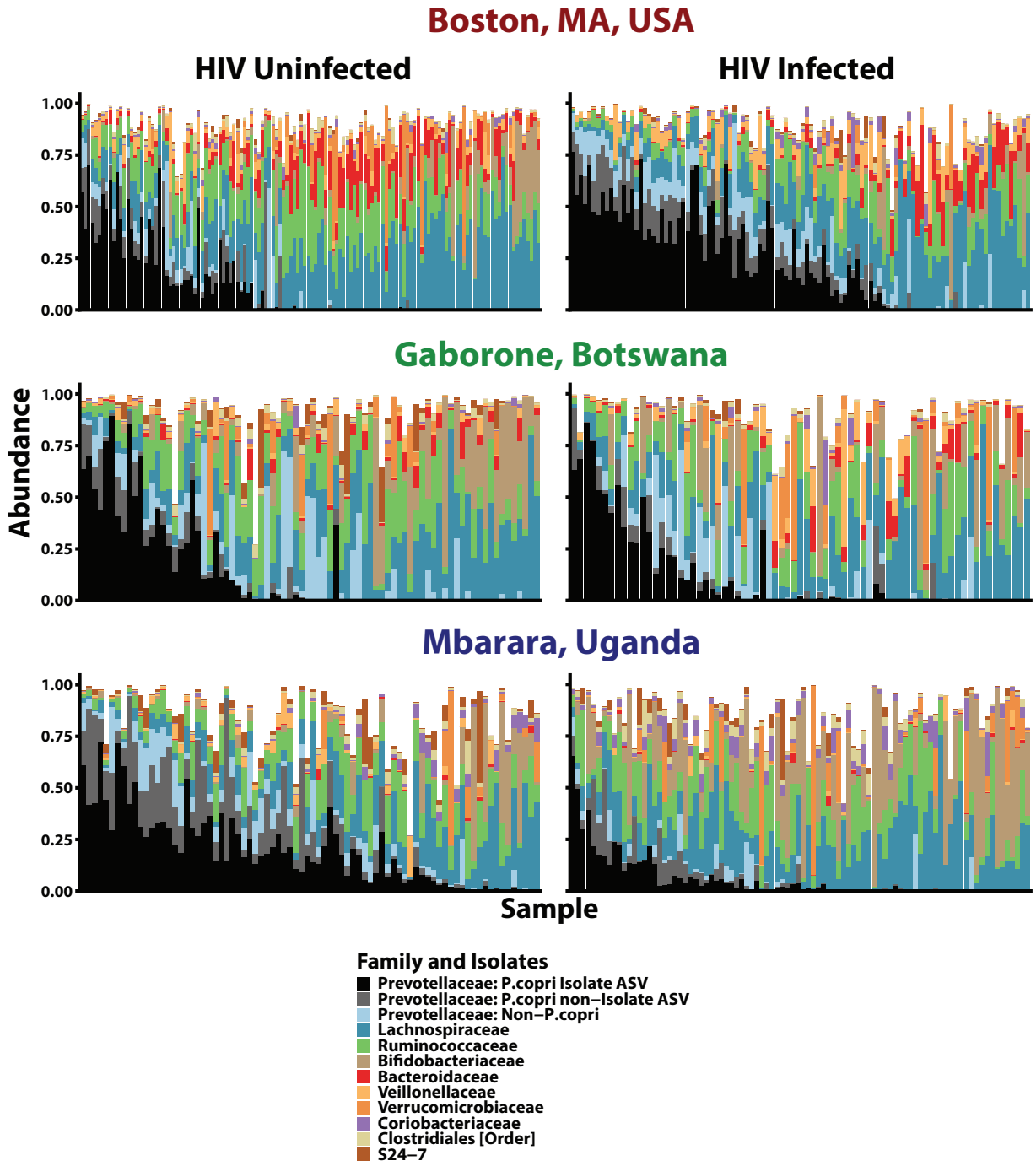

Figure 3.19. Fractional abundance composition of subjects' microbiomes at the family level with sub-division of Prevotellaceae. Subjects' samples are arrayed on the x-axis. Graphs in columns share the same HIV phenotype and graphs in rows share the same geographic cohort. Colors represent the 10 most abundant bacterial families with Prevotellaceae split into three groups representing Non-*copri Prevotella*, *P. copri* whose ASVs do not match an isolate, and *P. copri* whose ASVs do match an isolate. Reads that are not assigned to these families are left blank, meaning some bars will not reach the 1.00 fraction.

To collapse the isolate clones into unique genomes, we quantified pairwise dissimilarity with the binary jaccard distance calculated on the presence/absence of gene clusters created with 95% amino acid similarity (similar to the heat map in **Figure 3.20**). We determined that a pairwise distance of 0.40 maximized the number of collapsed clones without collapsing two clones from different samples. After collapsing, we had 29 unique genomes (17 from Boston, MA, USA; 12 from Mbarara, Uganda), which were hierarchically clustered using the complete linkage method (**Figure 3.20**). The clustering significantly separated the two geographical groups ($p < 0.002$) as calculated by a mantel test against groups, demonstrating that more genes are shared among isolates that were derived from the same geographic cohort. On the gene presence/absence heat map, a total core genome and a subclade core genome are visible as well as the accessory genomes that are often private or shared by only a few unique isolates.

In order to explore the functional genomic capacity of each unique isolate, we assigned KEGG orthology designations to the gene clusters, which could then be assembled into metabolic pathways and protein groups. From this, we could determine the completeness of these pathways and the presence or absence of functional proteins in each unique isolate. Using a combination of statistical screening and manual curation, we identified several KEGG pathways and functional groups whose completeness or functionality was significantly enriched or trending toward enrichment by geography, suggesting that the differences in genes seen across geographies by hierarchical clustering could lead to downstream functional differences.

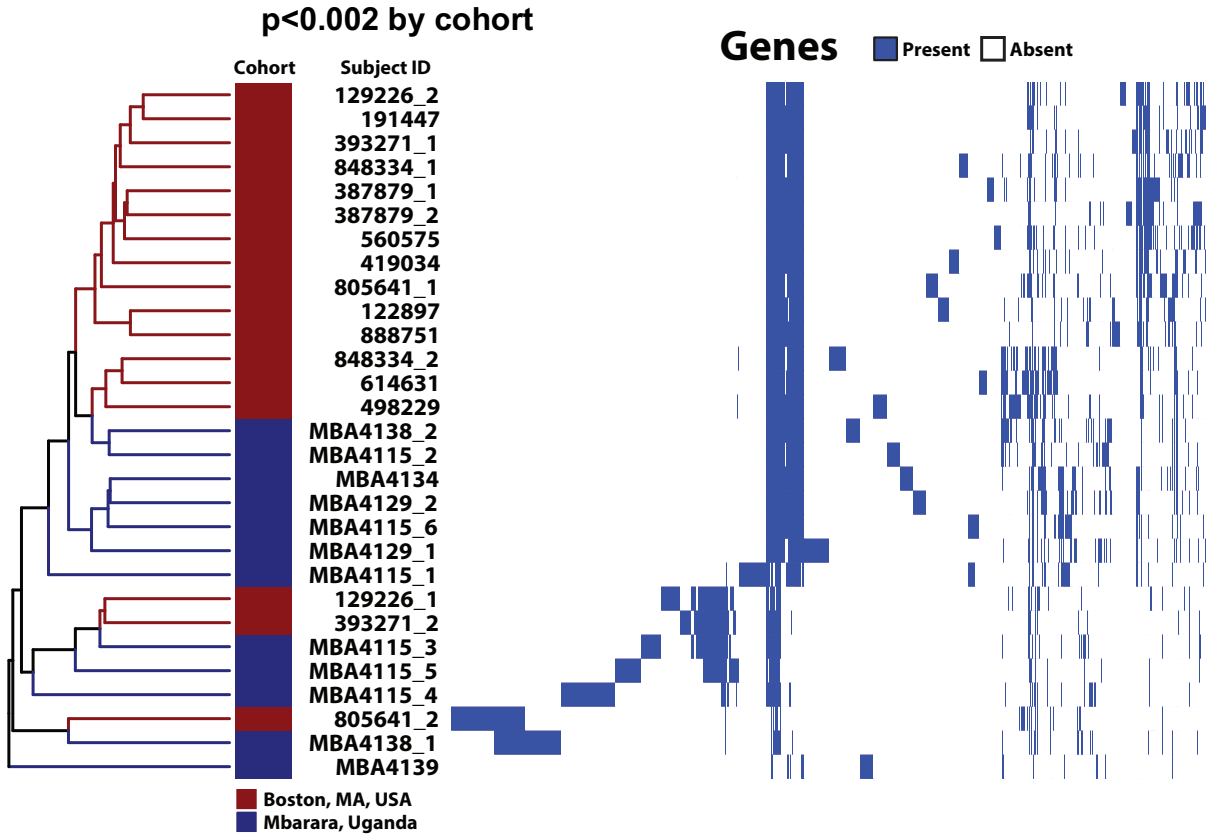

Figure 3.20. Gene content comparison and clustering of unique *P. copri* isolates. Isolate number and subject ID are arrayed on the y-axis with cohort membership. Genes are arrayed on the x-axis and heat map panel colors indicate presence or absence of a gene cluster in the corresponding genome. Genes are determined by clustering at 95% amino acid similarity. Dendrogram clades with membership from only one cohort are colored accordingly. All clustering was performed on binary jaccard distance calculated on gene cluster presence/absence tables. Rows (isolates) are hierarchically clustered using the complete linkage method. Columns (genes) are hierarchically clustered using Ward's method without squaring. P-value of clustering against cohort grouping calculated by Mantel Test against groups.

Among those pathways enriched or trending towards enrichment were several that would allow *P. copri* to respond to its environment or utilize available sources of energy. Some examples of these include two-component regulatory response systems, proteins involved in bacterial chemotaxis, and sugar hydrolases (**Figure 3.21**). The two histidine kinase-driven regulatory systems BaeS/BaeR and ResE/ResD were both enriched in the Ugandan isolates in comparison to the American *P. copri* ($p = .043$ for each). BaeS/BaeR systems often function to upregulate protective gene products such as efflux pumps in gram-negative bacteria in response to envelope stress, such as that induced by copper ions, flavonoids, indole, or antibiotics (Koler et al., 2016; Leblanc et al., 2011; Lin et al., 2014). The ResE/ResD family of proteins serve to drive nitrate respiration, biofilm formation, and the production of intermediates necessary for anaerobic growth in a low oxygen environment (Hartig et al., 2004; Nakano et al., 1996; Zhou et al., 2018a). In addition to the previously mentioned protein systems that enable environmental responsiveness, genes associated with the bacterial chemotaxis sensor MCP were trending toward enrichment in Uganda ($p = 0.083$). MCP and its downstream effectors serve to drive movement in response to external stimuli and are well represented in *Prevotella* within the oral cavity (Salah Ud-Din and Roujeinikova, 2017; Wang et al., 2013).

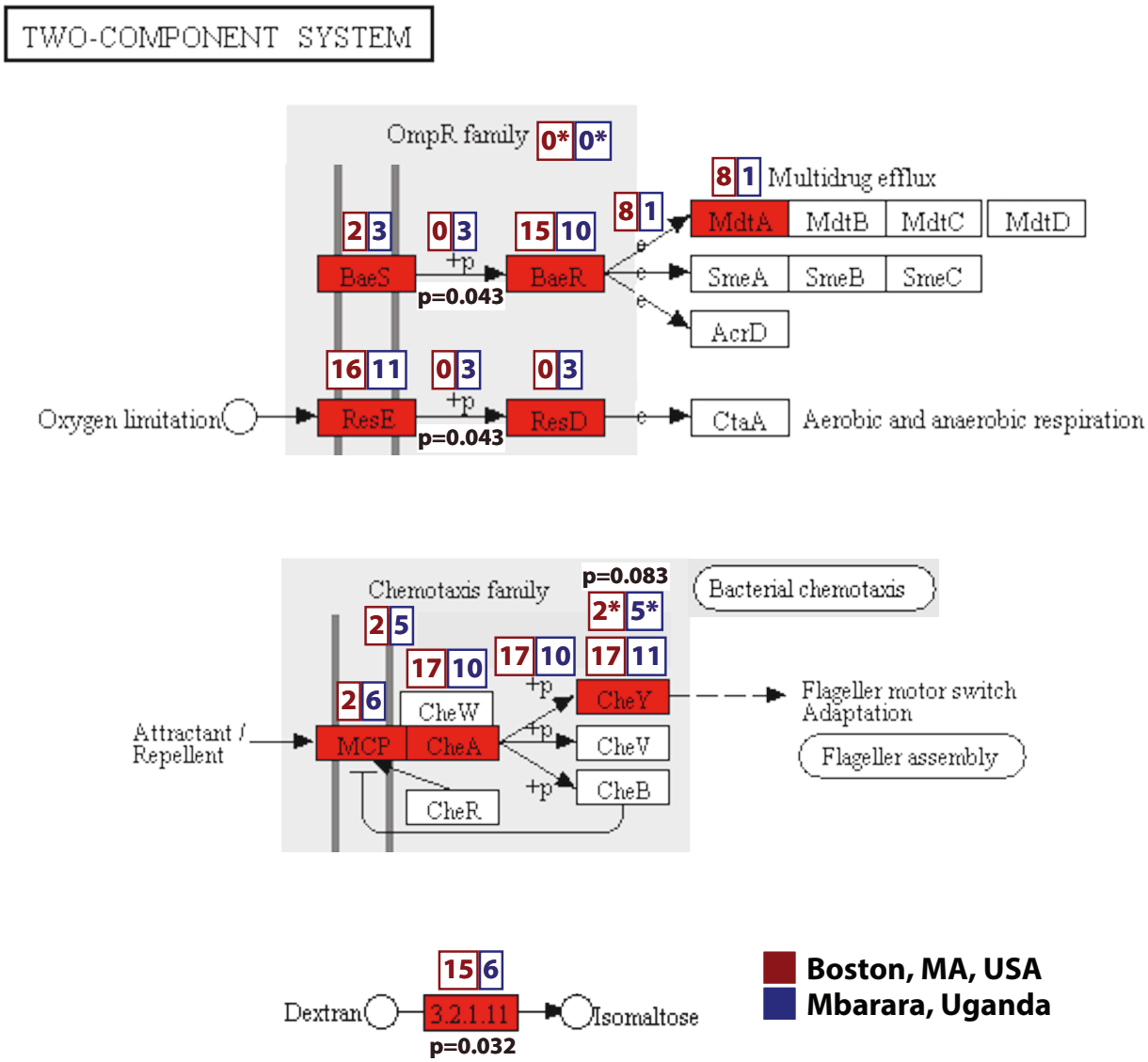

Figure 3.21. Select KEGG pathways enriched or trending towards enrichment by geography in *Prevotella Copri* isolates. A selection of four representative pathways reconstructed from assembled genomes of *P. copri* isolates. Enzymes on a red background are present in at least one isolate. Enzymes with a white background were not present in any isolate. Numbers in boxes directly above the red enzymes represent the number of isolates that had that enzyme and the numbers are split into counts of American (red) and Ugandan (blue) isolates. Numbers in boxes between two enzymes represent the number of isolates that had both enzymes. Boxes with asterisks are the numbers of isolates that had the complete pathway. P-values were calculated by Barnard's Unconditional test (a more powerful alternative to Fisher's exact test) using the total numbers of unique isolates in each location (Boston = 17, Mbarara = 12).

The catabolic enzyme dextran hydrolase is significantly enriched in *P. copri* isolates from Boston ($p = 0.032$). This enzyme hydrolyzes the glucose-based polysaccharide dextran to produce glucose disaccharides and is a potential source of energy for the microbiota (Sarbini et al., 2011). Similar to the MCP functional group, dextran hydrolase activity has been previously demonstrated in numerous oral *Prevotella* (Igarashi et al., 1998), so it is not unexpected that this enzyme would be abundant in *P. copri*. In addition, the substrate dextran is widely used as a food additive in roles such as a thickener for jam or ice cream (Bhavani and Nisha, 2010) that might result in a greater prevalence in HICs and consequently greater evolutionary pressure on the enteric microbiome of individuals in those geographies to adapt to use it as an energy source (Truong et al., 2017). The many functional differences between *P. copri* isolated from disparate geographic locations provides support for the conclusion that substantial and meaningful variation can exist at the sub-species level and could help explain the contrasting associations we observed in our study and those reported in the literature (Metwaly and Haller, 2019; Truong et al., 2017).

3.10 Discussion: Relationships between HIV infection, gut microbial community, and systemic inflammation are greatly influenced by host context

In assessing which ASVs were associated with differences in HIV infection status, the majority of the ASVs we found were differentially abundant in only one geographic cohort (**Figure 3.14**), implying that HIV-associated microbiota differences can vary greatly across geographic contexts. Even among those ASVs whose association with HIV was present in multiple cohorts, the directionality and magnitude of this association varied widely, with some ASVs exhibiting HIV-associated enrichment in one cohort and HIV-associated depletion in another. Of the 194 ASVs identified as differentially

abundant in any cohort, only 1 ASV exhibited the same directionality of HIV association in all three cohorts. In addition, within the Boston HIC cohort MSM behavior must also be taken into account when attempting to predict HIV-associated differences, as we observe that sexual behavior can act as an interacting factor that further modifies the microbial differences. These findings demonstrate that, although there were HIV-associated differences in gut bacterial communities in all three cohorts, the specific differences are particularly context-dependent when taking into account the expected baseline microbial community of an individual.

These observations fit into a larger model derived from our data, in which HIV infection drives unique microbial changes for each of the different baseline gut microbial communities present in individuals in different geographic locations. Within this context, the variations in associations between bacterial taxa and inflammatory markers suggest that HIV infection may disrupt some pre-existing anti-inflammatory microbial functions while also allowing the emergence of new relationships, though the nature of the changes in these relationships probably vary by the original microbial milieu. While it may seem trivial to conclude that HIV-associated differences are dependent on the starting microbial community, there are numerous examples in the literature of microbiome perturbations that are so strong and repeatable as to drive similar microbial changes. For example, extreme dietary interventions such as diets high in protein and fat and very low in carbohydrates have been shown to produce characteristic community endpoints regardless of the membership and structure of the starting microbial community (David et al., 2014). By showing that HIV infection has the characteristics of a context-dependent enteric microbiome perturbation, we provide

valuable insight that can aid further studies and therapeutic development within this field. This knowledge can help investigators avoid confounders, latent variables, and similar pitfalls that might arise when trying to conduct studies or deploy and measure interventions on an international scale.

3.11 Discussion: Limitations of this study

There were several limitations of this study. As this work was conducted at three sites spanning two continents and managed by three different groups with different study objectives, there was not always complete coordination across all three locations. While this did not have a major impact on the conclusions of this study, there were many categories of clinical metadata that were only collected at one or two study sites and therefore could not be compared between the three cohorts. This is an unfortunate lost opportunity as these were effectively partially collected patient data in the context of our study that combined all three sites. Had we been able to collect more metadata at all three sites, we would have been able to investigate far more possible hypotheses and characterize more relationships. One important example of this is with regards to dietary information. While we had dietary information for all three cohorts, these were collected with disparate methodologies and so we were unable to compare diet between these populations, which could have possibly provided insight into dietary factors driving geographical microbiome differences. In addition, only the dietary data from the Botswanan cohort was detailed and consistent enough to meaningfully analyze in comparison to the microbiome. The inability to do a complete analysis of the dietary contribution to our observed microbiome differences is unfortunate as diet has been

shown to have a substantial effect on the microbiome (David et al., 2014; Wu et al., 2011).

Another broad category of limitations to this study is possible confounders that we were not able to fully control for. For example, we did not have information regarding sexual practices such as MSM in either the Botswanan or Ugandan cohort, as MSM behavior can substantially affect the gut microbiome (Armstrong et al., 2018; Fulcher et al., 2018; Kehrmann et al., 2019; Kelley et al., 2017; Noguera-Julian et al., 2016; Nowak et al., 2019; Nowak et al., 2017; Pescatore et al., 2018). However, this is less concerning than a lack of this data would be in an HIC environment, as the HIV epidemic in sub-Saharan Africa is more characterized by heterosexual transmission of the virus (Fettig et al., 2014), meaning that the presence of MSM behavior is likely to be relatively less prevalent among our study subjects. Similarly, we did not have data on whether subjects in the Botswanan cohort lived in the urban environment of Gaborone or travelled from the more rural areas outside the city. As urbanization is associated with differences in the gut microbiome (Ayeni et al., 2018; Winglee et al., 2017), this may have affected our findings to some extent. Another possible confounder was that all the HIV-infected subjects in the Ugandan cohort were receiving co-trimoxazole prophylaxis which could have altered their microbiota. Co-trimoxazole is lightly studied compared to other antibiotics and the data is somewhat conflicting, but overall it may have a relatively more mild effect on the gut microbiota, as discussed in chapter 2 as well as other studies (de Bonnecaze et al., 2018; Hara et al., 2012; Kofteridis et al., 2004; Mavromanolakis et al., 1997; Monaco et al., 2016; Oldenburg et al., 2018; Stamey et al., 1977).

3.12 Discussion: Conclusion

Here we report one of the first studies examining the relationship between HIV infection and the enteric microbiome to encompass three diverse geographic locations on two continents that represent the many dimensions of the HIV epidemic and thereby allow the direct comparison of HIV-associated gut microbiome differences in HICs and LICs. We demonstrate that there are unique HIV-associated gut microbial differences, both at the whole-community and individual taxa levels, at each of these locations. These microbial differences also coincide with differences in the relationship between microbial taxa and host inflammatory markers. In addition, while there are members of the genus *Prevotella* found in all of these geographic locations, the functional associations of individual ASVs within this genus appear to shift substantially with differences in geographical context. Taken together, these findings suggest that HIV infection could alter anti-inflammatory microbial functions and allow the emergence of new functions, though all of these effects are highly context-dependent on the pre-existing gut microbial community. These changes may also occur at high taxonomic resolution, with functional variation and shifts possibly occurring at the species or strain level. This model has significant implications for efforts that intend to translate the results of HIV-microbiome studies between different geographical locations. Our findings suggest that in order to draw conclusions about the relationship between HIV infection, enteric microbiome changes, and host effects, studies must be conducted within the relevant patient populations, as inter-population variation might be so great as to drastically change conclusions.

Chapter 4: Significance and Future Directions

4.1 Significance

HIV is responsible for a substantial burden of disease on a global scale, with some of the greatest effects in LICs and sub-Saharan Africa (UNAIDS, 2018). With the development and deployment of effective ART, acute mortality from HIV as an infectious disease has declined, but chronic sequelae in infected individuals still result in significantly increased overall mortality for these patients (Collaboration of Observational et al., 2012; Egger et al., 2002; Katz and Maughan-Brown, 2017). As this excess mortality is closely linked with inflammation (Centers for Disease Control and Prevention, 2008; Egger et al., 2002; Lewden et al., 2007; Lohse et al., 2007; Triant et al., 2008), the gut microbiome is a potential contributor to this mortality by an inflammation-related mechanism. Supporting this model are the findings that HIV disrupts the enteric immune system that normally curates the gut microbiome (Brenchley et al., 2004; Douek et al., 2009; Klatt et al., 2008; Sandler and Douek, 2012) and that disrupted gut microbial communities are capable of driving host inflammation and pathology in many other contexts (Cho and Blaser, 2012; Friedrich, 2013; Littman and Pamer, 2011). While there has been a growing body of research investigating this potential connection between HIV infection and the gut microbiome (**Tables 1.5** and **1.6**), this work has largely been conducted in HICs rather than LICs where the impact of the HIV epidemic is greatest (**Figure 1.3**). This is particularly concerning in the context of gut microbiome research, as geographical differences are associated with some of the greatest variation in gut microbiome composition and structure (**Figure 1.4**).

To address this gap in prior research, we undertook a study to directly compare HIV-associated enteric microbiome differences in HICs and LICs. Our work, conducted on a background that is representative of the global span of gut microbiome

compositions (**Figure 3.1**), demonstrates that HIV-associated gut microbiome differences are highly dependent on host context, including such factors as geography (**Figure 3.6**) and sexual behaviors (**Figure 3.9**). These findings have great utility for informing further study and development of therapeutic interventions directed at the gut microbiome in HIV infection. With regards to further study, we have highlighted the significance of a potentially strong confounder that could interfere with the ability to detect true relationships between the microbiome and host health in HIV infection or introduce spurious relationships, similar to the complicating effect that MSM behavior has on the study of HIV infection and the gut microbiome (Armstrong et al., 2018; Fulcher et al., 2018; Kehrmann et al., 2019; Kelley et al., 2017; Noguera-Julian et al., 2016; Nowak et al., 2019; Nowak et al., 2017; Pescatore et al., 2018). Taking care to control for host context factors such as geography or MSM behavior will improve the efficiency and efficacy of future study of HIV and the microbiome.

In relation to the design of therapeutic interventions targeted at alleviating chronic inflammation, pathology, and mortality in HIV infection, our findings can contribute to the development of more effective therapies. Current trials of microbiome-directed therapies such as probiotics in HIV have produced conflicting results and experienced only limited success (Cunningham-Rundles et al., 2011; d'Ettorre et al., 2015; González-Hernández et al., 2012; Hummelen et al., 2011; Irvine et al., 2010; Miller et al., 2016; Monachese et al., 2011; Reid, 2010; Schunter et al., 2012; Stiksrud et al., 2015; Villar-Garcia et al., 2015). This is potentially due to a mismatch between the mechanism of the therapy and the healthy “baseline” community of the patients. This mismatch will be increased if these therapies are implemented more widely in the future to patient populations with

meaningfully different host contexts. For example, a therapeutic intervention developed in one host context, such as individuals in HICs, may have severely impaired effectiveness when deployed in another host context, such as individuals in LICs, that is significantly different with regard to baseline microbiota and its interaction with HIV infection. On a broader public health and health equity level, the importance of host context in microbiome-targeted therapeutics for this indication may serve to exacerbate the already-present inequality between HICs and LICs in HIV as it is likely that a substantial fraction of development of these interventions will be conducted in HICs, thereby limiting their application in LICs, especially in sub-Saharan Africa, that bear a disproportionate burden of HIV disease. The insight gained from our study can help inform and drive the development of microbiome-targeted therapeutics that are well-suited to the HIV-affected populations in LICs that might not otherwise benefit as greatly from these medical interventions.

Our work helps further characterize *Prevotella* and contributes to related methods development and validation for *Prevotella* isolation. *Prevotella* are an essential and important constituent of the gut microbiome, with associations with geography (De Filippo et al., 2010; Gomez et al., 2016; Gorvitovskaia et al., 2016; Lozupone et al., 2013; Martinez et al., 2015; Obregon-Tito et al., 2015; Schnorr et al., 2014; Smits et al., 2017; Wu et al., 2011; Yatsunenko et al., 2012), inflammatory diseases like rheumatoid arthritis (Scher et al., 2013) and osteolytic bone disease (Lukens et al., 2014), and diet (De Filippis et al., 2019; De Filippis et al., 2016; Kovatcheva-Datchary et al., 2015; Metwaly and Haller, 2019). As *P. copri* normally constitutes the overwhelming majority of *Prevotella* species in the gut (Franke and Deppenmeier, 2018; Human Microbiome

Project, 2012; Ley, 2016) and is a genomically dynamic organism that contributes greatly to geographically-associated microbiome differences (Vangay et al., 2018), this species in particular merits increased focus. However, since the original isolation of the *P. copri* type strain in Japan (Hayashi et al., 2007), there have only been 22 sequenced *P. copri* deposited in the Refseq database: 19 from Shenzhen, China and one each from the United Kingdom, India, and Japan (the type strain). A recent large metagenomic study analyzing over 1000 samples found that *P. copri* metagenomes could be divided into 4 geographically-associated subclades (Truong et al., 2017), implying that, similar to the gut microbiome as a whole, the greatest amount of variation within *P. copri* is associated with geographical differences. Taking into account the general importance of *P. copri*, its high abundance and prevalence in LICs, and the degree to which it varies geographically further highlights the discordance between the geographic distribution of resources for studying *P. copri* (as indicated by the origins of the sequenced genomes) and the locations where study of *P. copri* could have the greatest impact.

The scope of this project included the development and validation of techniques to efficiently isolate *P. copri* from clinically relevant cohorts. We successfully isolated an estimated 27 new *P. copri* variants from previously unrepresented populations, and validated our culturing strategy as a rapid and accurate phenotypic test for differentiating *P. copri* from other *Prevotella* and *Bacteroides* in order to accelerate isolation. By drastically expanding the pool of sequenced *P. copri* genomes, we are greatly advancing the study of *P. copri* at sub-species or strain level. In addition to directly allowing the discovery of strain-specific differences between our *P. copri*

isolates, the genomes will aid in the further single genome or metagenomic assembly of new *P. copri* genomes. Interrogating the strain-level variation within *P. copri* is a particularly fruitful scientific pursuit, because recent work has suggested that strain-level variation might be the most important scale at which to examine organisms in the *Prevotella* genus and that the association between *Prevotella* and host diet may be linked with strain-level diversity (De Filippis et al., 2019; Metwaly and Haller, 2019).

Our promising results in identifying geographically associated differences within *P. copri* isolates also serve to provide greater motivation for exploring *Prevotella* as well as important enteric bacteria in general at a more global scale. This conclusion regarding the necessity of exploring the human-associated microbiota on a global scale is applicable both to our specific field of HIV-gut microbiome interactions as well as the broader discipline of microbiome study, and is in concordance with other work showing the functionally relevant degree to which the enteric microbiome can vary across the world (Smits et al., 2017). The findings from our comparative genomic investigation of *P. copri* genomes are preliminary, but suggest that within *P. copri*, meaningful geographic variation could include genes coding for environmental response elements, chemotaxis, and catabolic enzymes (**Figure 3.21**). This variation could allow different strains of *P. copri* to be better suited for their unique environment by modifying the repertoire of environmental changes to which they can respond, modifying their response to the same set of environmental changes, or allowing them to utilize different sources of nutrients. Lastly, our findings also support the importance of isolate culturing and sequencing as complementary method to strictly metagenomic sequencing, as they

provide a unique insight into genomic variation that is guaranteed to truly exist between two different bacterial strains.

Beyond the field of HIV-microbiome associations and Prevotella genomics, the work described in this dissertation also contributes to the fields of MSM-associated microbiome, diet and the microbiome, and urbanization and the microbiome. We have assembled a large cohort containing numerous subjects representing the combinations of MSM behavior and HIV phenotype. Our data add to the body of work describing an MSM-characteristic microbiome (Section 1.7) as well as the HIV-associated differences among MSM (Section 3.5). With regards to diet and the microbiome, we report that in the mixed urban-rural population of Gaborone in the LIC of Botswana, we did not detect a nutrient-associated signal within the microbiota. This finding adds to the growing body of knowledge concerning the influence of diet on the microbiome under different contexts and dietary variation. Because our study included two sites in sub-Saharan Africa, one rural and one urban, and found differences between each of the sites (**Figure 3.1**), our data support the prevailing model that even within similar geographic regions, the degree of urbanization can have a significant effect on the microbiome (Ayeni et al., 2018).

In a more abstract sense, our study contributes to the body of literature describing the general effect size and universality of gut microbiome perturbations, further elucidates the characteristics that meaningfully differentiate these perturbations, and helps classify the specific perturbation of HIV infection within this intellectual framework. The data presented here will aid in defining the boundaries and distribution of the effect sizes of gut microbiome perturbations that span from those with a great

effect on the gut microbiome, such as drastic diet changes or significant gastrointestinal infections, to those with an inconsequential or spurious effect on the gut microbiome, such as choice of eyewear or television viewing habits. By accumulating more data with regards to which perturbations in general are significant or trivial, we also help create a kind of informal classifier composed of characteristics and relationships that will enhance our ability to *a priori* predict the effect of undescribed perturbations. For example, if there were no prior data on what the introduction of a selected microorganism would have on the gut microbiota, the effect of that addition might be predicted by the relative balance of similarities between that organism and a pathogen such as *Salmonella enterica* or that organism and a commensal bacterium. Lastly, in a more concrete sense our study helps situate HIV infection as a microbiome perturbation within this intellectual framework, providing support for the conclusion that HIV infection can have a substantial effect on the gut microbiome, though not on the scale of previously described major perturbations such as severe gastrointestinal infection.

Similar to the contribution to general knowledge regarding the effect size of microbiome perturbations, our work helps elaborate on a related dimension of universality or variance. Our findings can aid in the definition of paradigms surrounding the universality or variance of microbiome perturbations that are comparable to the intellectual framework surrounding microbiome perturbation effect size described above. As a concrete example in our case, our data suggest that HIV infection is a high variance microbiome perturbation that does not produce universal results. This is due to the great influence that host context and other factors have on the magnitude and directionality of gut microbiome change during HIV infection, with substantially different

outcomes resulting when this perturbation is applied to different starting microbial communities and host contexts. This is in contrast to a microbiome perturbation that has a more universal effectiveness and outcome that is less dependent on outside factors. A previously referenced example of a perturbation with these characteristics is drastic diet interventions, whose effect on the gut microbiome is so strong that its outcome is significant and consistent regardless of pre-existing host state (David et al., 2014).

As research into the microbiome is progressing towards human health applications, the concepts above can be further illuminated and contextualized with the use of clarifying analogies in medicine. In the larger context of human disease, there are diseases whose effects are so robust that outcomes are consistent and there are also diseases whose outcomes vary wildly, often as the result of an interaction with a host or environmental factor. Examples of diseases or health events with robust effects are those such as severe intracranial hemorrhages or aggressive metastatic neoplasms which are highly likely to result in disastrous health outcomes and small cuts or abrasions which are unlikely to lead to a severe health outcome. Conditions with state-dependent effects include many autoimmune diseases such as asthma or type 1 diabetes mellitus, which often involve predispositions due to host factors but often require an environmental stimulus to trigger. Depending on the degree of interaction from that external stimulus, the course of the disease can vary widely. It is important to distinguish, recognize, and increase our knowledge about differences such as this in medicine, study of the microbiome, and the overlap between these two fields, as differently robust diseases and perturbations demand a fundamentally different responses and conceptual frameworks. This can include the importance and urgency of

an intervention, as the course of less robust diseases and microbiome perturbations can be more greatly modified than those of more robust diseases, and therefore interventions can have greater value in less robust perturbations when viewed in this light.

4.2 Future Directions

The work presented here could be extended by a wide variety of future investigation that further explores the microbial communities of interest as well as extends into more functional characterizations of individual bacterial isolates. While a substantial fraction of the *P. copri* ASVs observed in our original 16S rRNA gene sequencing were accounted for by the isolates we obtained (**Figure 3.19**), there were still samples that had relatively low coverage of these ASVs by our isolates. Additional culturing from remaining fecal samples could achieve much more complete representation of the full *P. copri* biodiversity contained within our samples. These efforts to expand the pool of isolates to target unrepresented ASVs could include screening steps with Sanger sequencing of 16S rRNA gene amplicons in order to apply the entire isolation pipeline to only those individual colonies that contain unrepresented 16S rRNA gene sequences.

In addition to expanding the collection of *P. copri* isolates, another important next step in this work is functional testing and characterization of P*. copri* isolates. Based on preliminary comparative genomic findings of possible functional differences (**Figure 3.21**), directed testing of the ability of the isolates to catabolize dextran and use it as an energy source could be informative. To complement directed metabolic testing, we could use broad functional screening systems such as Biolog, which would expose the

bacteria to a wide variety of metabolites and inhibitors in order to more fully characterize their functional phenotype. This testing and screening could also be performed in defined mixed microbial communities containing combinations of the *P. copri* isolates as well as other bacterial isolates in order to probe the ability of these organisms to complement one another and form complete metabolic pathways on a community scale. Learning more about the *P. copri* isolates would greatly increase their value as a reagent to use in future experimentation.

Another next step in the continuation of this work would be the use of additional sequencing-based tools to characterize the microbial communities in the fecal samples collected from our subjects. We originally used amplicon sequencing of the 16S rRNA gene to quantify the gut microbiota of the subjects in our study due to the large study size and, at the time, the substantial resource outlay necessary for other sequencing strategies. However, since then other sequencing methodologies have become more accessible and scalable. Shotgun high-throughput sequencing of the total metagenomic content of a bacterial community can offer several advantages over 16S rRNA gene amplicon sequencing, especially in fecal samples where bacterial DNA is abundant relative to host DNA (Marotz et al., 2018). Available analysis tools can use metagenomic data directly to characterize and quantify bacterial taxonomy, often at greater resolution than that possible with 16S rRNA gene amplicon sequencing (Segata et al., 2012), making strain-level identification of microbes much more possible (Truong et al., 2017). There are also computational techniques for reconstructing metabolic pathways present in metagenomic data (Franzosa et al., 2018). A newly emerging set of bioinformatics tools has greatly improved the ability to assemble genomes from

metagenomic sequencing (Nayfach et al., 2019; Pasolli et al., 2019), so shotgun sequencing of our samples would allow us to assemble genomes both to verify them against our isolate genomes and to discover new bacterial genomes.

The previously described sequencing approaches only probe the DNA of the bacterial portion of the gut microbiome. Future work could marry this with sequencing of the fungal (mycobiome) or viral (virome) components of the gut microbiome. We have already collaborated with investigators to perform virome work (Monaco et al., 2016), so enlarging our research to encompass other domains of life is a viable next step. In addition to expanding the taxonomic focus of our sequencing, we could also expand the nucleic acid focus of our sequencing by including RNA in order to determine bacterial transcriptomics. While genomic DNA only gives a window into the genetic potential of an organism (*i.e.* what it is capable of doing), sequencing RNA is the first step into characterizing what an organism is actually doing. We could perform a targeted transcriptomic investigation of those differentially abundant genes between America and Uganda in order to determine if the genomic differences are borne out in a transcriptomic context. Going beyond nucleic acid sequencing, we could also perform proteomics or metagenomics to profile the proteins and small molecules being produced within the bacterial communities. This would provide insight even further downstream in the translation of genotype into phenotype.

Most of the future work proposed so far is of a much more observational nature, rather than involving interventions or perturbations that more directly probe causality. With the resources of viable fecal samples and bacterial isolates, we could employ an *in vivo* mouse model to demonstrate a causative link between certain microbes of interest

and chronic inflammation and inflammation-related pathologies that are common to HIV infection. For example, cardiovascular disease (CVD) is one of the major long-term pathologies in HIV-infected individuals (**Table 1.4**) and there exists an established ApoE$^{-/-}$ C57BL/6 mouse model for assessing the effect of the microbiome on CVD phenotypes (Bennett et al., 2015; Fisher et al., 2012; Gregory et al., 2015; Koeth et al., 2013; Wang et al., 2011; Wang et al., 2015). These mice lack the cholesterol particle constituent apolipoprotein E and more rapidly develop atherosclerosis with lesions that resemble those found in humans (Meir and Leitersdorf, 2004; Nakashima et al., 1994). This model has previously been used to show that transplanted bacterial communities can modulate CVD disease status, even without a germ-free recipient (Gregory et al., 2015). We could perform *in vivo* transplantation of microbes of interest into this model to determine if HIV-associated microbial communities have the potential to induce CVD.

For an *in vivo* experiment using the ApoE$^{-/-}$ C57BL/6 mouse model, we could take advantage of the *P. copri* isolates we have collected by establishing them in the mice either singly, in combination, or as a supplement to a baseline community made from whole stool. Prior to transplant, the mice would receive a 3-week conditioning regimen with the antibiotics vancomycin, neomycin, metronidazole, and ampicillin to prepare them for transplant and at then end of this regimen, the animals would be gavaged on a schedule described in the literature (**Figure 4.1**) (Gregory et al., 2015). We could assess serum cardio-active metabolites such as TMAO and inflammatory serum cytokines IL-1β, IL-6, IL-8, TNF-α, and sCD14 (Brenchley and Douek, 2012; Brenchley et al., 2006; Federico et al., 2009; Haller et al., 2000; Jiang et al., 2009; Kim et al., 2016; Reimund et al., 1996; Rogler and Andus, 1998; Tanoue et al., 2008; Van

De Walle et al., 2010), which have all been previously referenced and could indicate an immune-related mechanism driving CVD pathology.

We could assess the degree of CVD pathology that develops within the transplanted mice as measured by aortic plaque, lipid-laden monocyte foam cells, and macrophage infiltration into vascular plaques (**Figure 4.2**). To quantify lipid content and foam cell abundance, isolated monocytes would be fixed and stained with Oil-Red-O and hematoxylin using protocols established in our lab similar to those previously published (Wang et al., 2011; Wang et al., 2015). Fixed cardiac tissue could be sectioned and then stained with Oil-Red-O and hematoxylin to identify atherosclerotic lesions or with immunohistochemical staining for Iba1 (ionized calcium-binding adapter molecule 1) to identify macrophages embedded in the vessel walls (Gregory et al., 2015). On these macrophages, we could measure surface expression of scavenger receptors by flow cytometry using antibodies for scavenger receptors SR-A1 and CD36 and macrophage markers F4/80 and CD11b (Cailhier et al., 2005; Kunjathoor et al., 2002; Schledzewski et al., 2006; Wang et al., 2011). By assessing these multiple markers of systemic inflammation and CVD pathology in this ApoE$^{-/-}$ C57BL/6 microbiome transplant mouse model, we could more strongly establish a link between specific HIV-associated (either by enrichment or depletion) microbes (such as our *P. copri* isolates) and the chronic complications of HIV that have been observed in population studies.

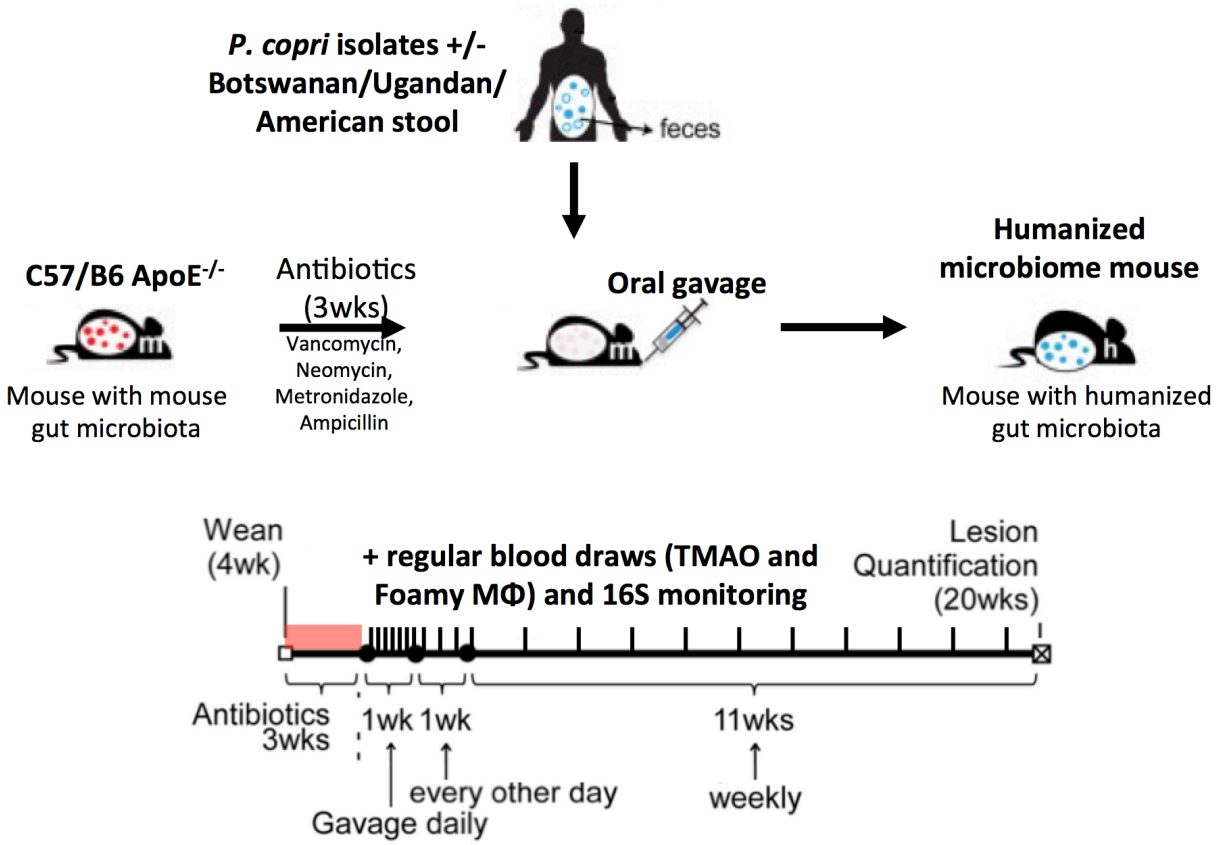

Figure 4.1. Schema for fecal transplant into ApoE$^{-/-}$ C57BL/6 mice to quantify induced inflammation. Conventionally raised mice are conditioned with antibiotic treatment and then orally gavaged with preparations of host stool and isolated bacteria as indicated in the figure. Regular blood draws and stool 16S rRNA gene amplicon sequencing is performed to monitor disease progression and fecal transplant engraftment. At 20 weeks mice are sacrificed in order to perform assays in **Figure 4.2**. Adapted from (Gregory et al., 2015).

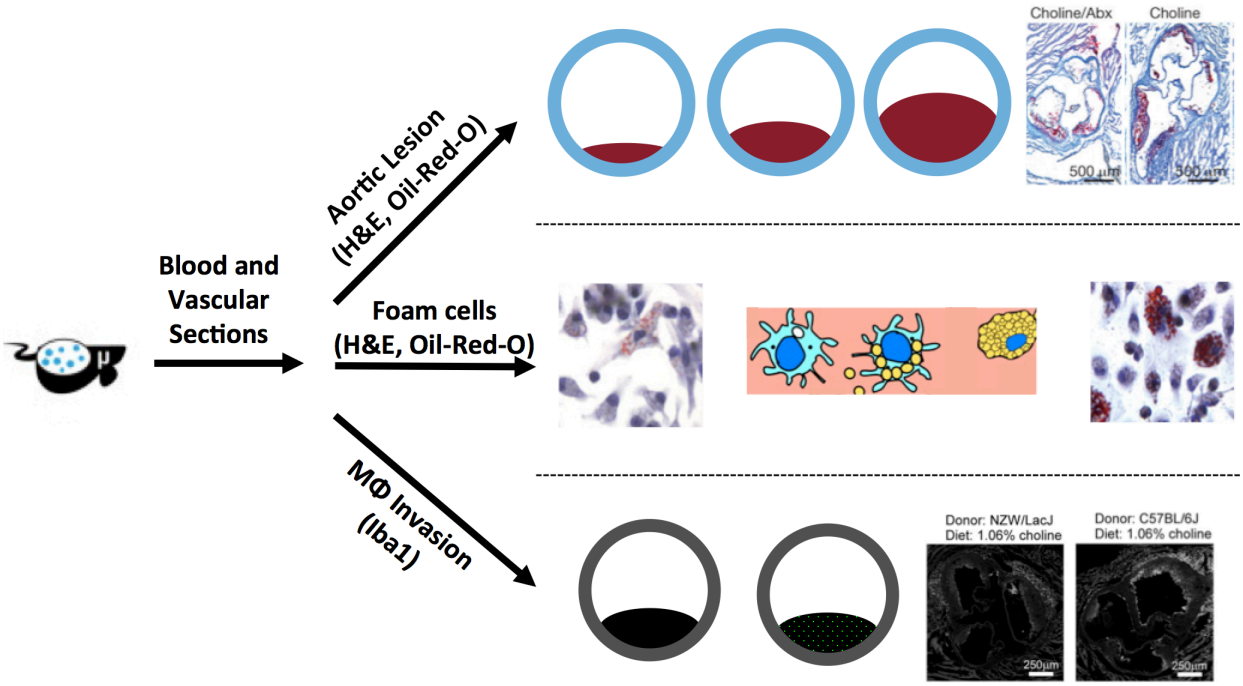

Figure 4.2. Cardiovascular disease assays performed on ApoE$^{-/-}$ C57BL/6 mice after sacrifice. Assays performed after sacrifice noted in **Figure 4.1**. Blood and vascular sections are taken from the mouse and stained with Oil-Red-O and Iba1 antibodies. Oil-Red-O staining of aortic cross sections is used to quantify the area of aortic lesions, Oil-Red-O staining of blood is used to quantify the abundance and size of lipid-laden monocyte foam cells, and Iba1 staining of aortic cross sections is used to quantify macrophage infiltration into aortic plaques. Adapted from (Gregory et al., 2015).

In addition to the previously described experiments that would serve to further the work described in this dissertation, there are several improvements over our current human study that could be implemented if we conducted additional observation human studies to more thoroughly probe the connections between HIV infection, the gut microbiota, and chronic inflammation and related pathologies. We could add a longitudinal component to the human studies either in a newly established cohort or by requesting that the subjects we previously collected from return to provide a second fecal sample. This would provide greater information about the relative stabilities of the microbial communities of HIV-uninfected and -infected individuals. Collection of consistent and standardized diet data would allow the comparison of diet between each of the cohort locations and could be used to determine what proportion of the differences between each geography could be attributed to diet. Diet information is also particularly important because, as stated before, strain-level diversity of *Prevotella* has been shown to be shaped by host diet (De Filippis et al., 2019; Metwaly and Haller, 2019) and diet can have a substantial effect on the gut microbiome (David et al., 2014; Wu et al., 2011), so having this information would allow us to account for diet-driven effects. Sampling methods allowing for the investigation of other microbial subpopulations of importance within the gut, such as mucosal swabs or biopsies for mucosal-associated bacteria, would also be a valuable addition to future human studies, as mucosal-associated bacteria might have a greater effect on the gut-associated immune cells with which they are in close contact. We could also collect additional host data, such as intestinal permeability measured by the lactulose and mannitol challenge test that would help us determine intermediate phenomena and mechanisms that might

be involved in the linkage between the gut microbiota and inflammation in HIV. In a more ambitious undertaking, we could design a study to explore true “changes” in the gut microbiome caused by HIV, rather than the “differences” we observed here. This would involve sampling individuals before and after HIV acquisition, possibly by following a cohort of high-risk HIV-uninfected subjects and sampling them regularly especially if they acquire HIV. While further human studies would be a substantial undertaking, they would provide great insight into the relationships explored in this dissertation.

Chapter 5: Materials and Methods

5.1 First Ugandan study cohort (Chapter 2)

Matched, de-identified stool and plasma samples were collected from 122 subjects enrolled from the Mbarara Regional Referral Hospital in Uganda as approved by the Institutional Review Boards at the Mbarara University of Science and Technology, Ugandan National Council of Science and Technology, and Partners Healthcare. All participants gave written informed consent. This cohort was comprised of 42 subjects with untreated HIV disease, 40 location-matched samples from subjects on long-term ART therapy (>5 years) and 40 HIV-uninfected subjects. Subjects presenting to the HIV clinic for HIV testing were recruited into either the HIV-positive untreated arm or the HIV-negative arm depending on HIV test results. Subjects with other comorbidities were not excluded. Data collected included demographics, vital signs, medication history including antibiotic use, clinical symptoms, HIV RNA, CD4 T cell counts at stool collection, CD4 at ART initiation, water source, food security, farming, and other laboratory results (**Table 2.1**). Stool samples were collected in RNAlater and frozen at -80°C. Plasma samples were collected in acid citrate dextrose tubes and frozen at -80°C. Samples were shipped on dry ice and complied with the Material Transfer Agreement between Uganda, MGH and Washington University.

5.2 Bacterial 16S rRNA analysis (Chapter 2)

5.2.1 Stool pulverization (Chapter 2)

Aliquots of pulverized human stool (100-200mg) were processed as previously described (Reyes et al., 2013). Briefly, stool was chipped from RNAlater on liquid nitrogen, samples were pulverized, aliquotted (approximately 200mg stool each) into 2-3 separate 2mL collections tubes (Sarstedt) and stored at -80°C until use. Aliquots were

used for total nucleic acid (TNA) extraction and VLP preparation, ensuring that similar parts of the stool samples were used for both extraction methods.

5.2.2 Human stool total nucleic acid extraction

Stool TNA was extracted from aliquots of pulverized human stool (~200mg) as previously described (Reyes et al., 2013) with modification. Briefly, 200µl of 1 mm diameter zirconia/silica beads (Biospec) were added to individual pulverized stool aliquots. 500µL of phenol:chloroform:isoamyl alcohol (Fisher Scientific, 25:24:1, pH 8.0), 500µL of 0.2µm-filtered 2x Buffer A (200mM NaCl, 200mM Tris, 20mM EDTA), and 210µL of 20% SDS were added to each sample. Samples were chilled on ice and homogenized using the highest setting on a BioSpec Mini-Beadbeater for 2 minutes at 4ºC. The homogenized samples were then centrifuged at 4ºC for 3 minutes at 7000 x *g*, and the aqueous phase was transferred to a clean tube. An equal volume of phenol:chloroform:isoamyl alcohol was added and mixed by vortexing. Samples were centrifuged at 16,000 x *g* for 5 minutes at room temperature and the aqueous phase transferred to a clean tube. Nucleic acid was precipitated with isopropanol and 3M sodium acetate (pH 5.5, Ambion) at -80°C for 20 minutes, then spun at maximum speed at 4°C for 30 minutes. The pellet was washed with 500µl 100% ethanol, centrifuged at 16,000 x *g* for 15 minutes at 4ºC, dried, and resuspended in 200ul of molecular grade Tris-EDTA buffer (Ambion). DNA was isolated from the total nucleic acid preparation using an AllPrep DNA/RNA Mini Kit (Qiagen) according to the manufacturer's instructions. Nine samples resulted in insufficient quantity of DNA for 16S studies (Chapter 2).

5.2.3 16S rRNA amplification and sequencing (Chapter 2)

Primer selection and polymerase chain reaction was performed as described previously (Caporaso et al., 2011). Briefly, each sample was amplified in triplicate, pooled, and confirmed by gel electrophoresis. PCR reactions contained 2.5μL 10X High Fidelity PCR Buffer (Invitrogen), 18.8μL RNase/DNase-free water, 0.5μL 10 mM dNTPs, 1μL 50mM MgSO4, 0.5μL each of the forward and reverse Golay-barcoded primers specific for the V4 region (515F/806R, 10μM final concentration), 0.1μL Platinum High Fidelity Taq (Invitrogen) and 3μL extracted total nucleic acid. Reactions were held at 94°C for 2 minutes to denature the DNA, with amplification for 26 cycles of 94°C for 15s, 50°C for 30s, and 68°C for 30s; a final extension of 2 minutes at 68°C (to ensure complete amplification). Amplicons were pooled and purified using 0.6x Agencourt AMPure XP beads (Beckman-Coulter) according to the manufacturer's instructions. The final pooled samples were sequenced on the Illumina MiSeq platform (Washington University Center for Genome Sciences; 2x250 standard run) in two separate runs.

5.2.416S rRNA amplification and sequencing (Chapter 3)

Primer selection and polymerase chain reaction was performed as described previously (Caporaso et al., 2011) with modifications from (Anahtar et al., 2016). Briefly, each sample was amplified in triplicate, pooled, and confirmed by gel electrophoresis. PCR reactions contained 5μL 5X Q5 Reaction Buffer (NEB), 16.25μL RNase/DNase-free water, 0.5μL 10 mM dNTPs, 0.5μL each of the forward and reverse Golay-barcoded primers specific for the V4 region (515F/806R, 10μM final concentration), 0.25μL Q5 Polymerase (NEB) and 2μL extracted total nucleic acid. Reactions were held at 98°C for 30 seconds to denature the DNA, with amplification for 30 cycles of 98°C for

150, 60°C for 30s, and 72°C for 20s; a final extension of 2 minutes at 72°C (to ensure complete amplification). Amplicons were pooled and the final pooled samples were sequenced on the Illumina MiSeq platform (Ragon Institute of MGH, MIT, and Harvard; 1x300 standard run).

5.2.5 16S rRNA analysis (Chapter 2)

Analysis of R1 16S sequence data was performed using QIIME (Quantitative Insights Into Microbial Ecology, version 1.9.1) (Caporaso et al., 2010b). Raw sequence fastq files were quality filtered and demultiplexed using default parameters with the following exceptions: PHRED quality score cut-off at 20, and reverse-complement mapping barcodes were used. Closed reference operational taxonomic units (OTUs) sharing 97% identity were clustered using the UCLUST algorithm (Edgar, 2010) and assigned taxonomy according to the Greengenes database (version 13.8) (McDonald et al., 2012). To standardize differences in the number of OTUs between sequencing runs, all samples were rarefied to 5000 OTUs (10 iterations without replacement; maximum of 5000 OTUs per sample; 10 rarefaction steps) and the relative number of sequences assigned to each OTU was calculated for each sample. Two samples did not achieve high enough OTUs for downstream analysis. Alpha diversity analysis was performed on rarefied data. Faith's phylogenetic diversity (Faith and Baker, 2006) and the Chao1 richness metric were calculated for all ten rarefied tables. Statistical analysis between groups was performed using the compare_alpha_diversity.py function of QIIME. Species accumulation rarefactions plots were determined using the specaccum function of the vegan R package (Oksanen et al., 2013). Beta-diversity was determined in Phyloseq (v1.10.0) (McMurdie and Holmes, 2012) using weighted UniFrac distances.

Differential abundance of bacterial taxa between experimental groups was determined using the PhyloSeq DESeq2 extension using the Wald significance test and a parametric fit type (v.1.6.3) (Anders and Huber, 2010; McMurdie and Holmes, 2012).

5.2.6 16S rRNA analysis (Chapter 3)

Analysis of R1 16S rRNA gene amplicon sequence data was performed using QIIME (Quantitative Insights Into Microbial Ecology, version 1.9.1) (Caporaso et al., 2010b) and DADA2 (Callahan et al., 2016). Raw sequence fastq files were quality filtered and demultiplexed with QIIME with filtering turned off: PHRED quality score cut-off at 0, and reverse-complement mapping barcodes were used. Reads were then assigned to ASVs using DADA2. Faith's phylogenetic diversity (Faith and Baker, 2006) and the Chao1 richness metric were calculated for all samples. PhyloSeq (v1.26.1) (McMurdie and Holmes, 2012) was used to calculate alpha and beta diversity metrics and to perform principal coordinate analysis.

5.3 sCD14 measurements in plasma (Chapter 2)

Plasma was collected in acid citrate dextrose tubes and aliquots underwent a single freeze-thaw cycle before analysis. sCD14 concentration was measured by ELISA (R&D Systems Human sCD14 Quantikine ELISA kit #DC140). Samples were thawed on ice, centrifuged for 10 minutes at 1,000 x *g* and 4°C, and the supernatant was diluted 200-fold in Calibrator Diluent RD5P (1X, R&D Systems) per the manufacturer's instructions (10 µL sample + 1990 µL Calibrator Diluent RD5P). Samples were tested in duplicate with standards on every plate. sCD14 concentration was determined from optical density (O.D.) measurements by subtracting the average zero standard O.D. and performing a log-log transformation of the standards in order to fit a linear regression.

5.4 Plasma inflammatory marker measurement (Chapter 3)

Biomarker quantification provided by University of Vermont (UVM) Laboratory for Clinical Biochemistry Research included sCD14, sCD163, FABP2 by sandwich ELISA (DC140, DC1630, DFBP20; R&D Systems) and IL-6 by MSD (K151QXG); FABP2 and sCD163 were performed in duplicate.

5.5 Oligotyping

We performed oligotyping analyses (Eren et al., 2011) on differentially abundant 16S V4 sequencing reads assigned by QIIME to the *Ruminococcus* genera or *Enterobacteriaceae* family that were not previously resolved at the species level. Sequences shorter than the indicated length when trimmed to Phred score >30 were removed before analysis to prevent excessive variation due to sequencing error. Representative sequences for each oligotype were searched in the BLAST nr/nt database. The following table details the parameters for each performance of oligotyping (**Table 5.1**). Representative sequences for each oligotype are indicated below.

Table 5.1: Parameters used for oligotyping.

Oligotyping Group	Minimum Read Length	Total Full-Length Reads	Reads Assigned to Oligotypes	Minimum Sequences per Oligotype (-A parameter)	Oligotype Base Locations of Interest	Taxa assigned
PhyloSeq, *Ruminococcus* sp. enriched in CD4 >200 vs. CD4 <200	230	191993	134162 (0.699)	2000	25 Bases: 0, 8, 9, 12, 57, 58, 68, 79, 94, 95, 98, 112, 158, 174, 177, 178, 181, 201, 212, 229, 232, 236, 237, 242, 249	*R. bromii* *R. callidus*
PhyloSeq, *Ruminococcus* sp. enriched in HIV-negative vs. CD4 <200	0	109848	80909 (0.737)	2000	19 Bases: 0, 9, 77, 98, 148, 158, 174, 178, 181, 201, 212, 225, 229, 232, 236, 237, 238, 242, 249	*R. bromii*

MaAsLin, *Enterobacteriaceae* enriched in CD4 <200	240	111936	100009 (0.893)	1000	15 Bases: 0, 28, 57, 113, 138, 183, 189, 223, 226, 228, 232, 238, 242, 243, 249	*Shigella* sp. or a closely-related *Escherichia* sp.

PhyloSeq, *Ruminococceae Ruminococcus* sp. enriched in CD4 >200 vs. CD4 <200

>Oligotype TGAATTAAAATTTGGAACTAACGTC
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGG
CTTTGCAAGTCAGATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACT
GTAGAGCTTGAGTGAAGTAGAGGCAGGCGGAATTCCCCGTGTAGCGGTGAAATGC
GTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCTTTAACTGA
CGCTGAGGCACGAAAGCGTGGGTAGCAAAC

> Oligotype TGAGTGAAAGTTTCGAACTTACGGC
TACGTAGGGAGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGTAGGCG
GGATGGCAAGTCAGATGTGAAAACTATGGGCTCAACCCATAGACTGCATTTGAAAC
TGTTGTTCTTGAGTGAGGTAGAGGTAAGCGGAATTCCTGGTGTAGCGGTGAAATG
CGTAGAGATCAGGAGGAACATCGGTGGCGAAGGCGGCTTACTGGGCCTTTACTGA
CGCTGAGGCTCGAAAGCGTGGGGAGCAAAC

> Oligotype TGAGCGAAAGTTTCGAACTTACGGC
TACGTAGGGAGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGTAGGCG
GGACGGCAAGTCAGATGTGAAAACTATGGGCTCAACCCATAGACTGCATTTGAAAC
TGTTGTTCTTGAGTGAGGTAGAGGTAAGCGGAATTCCTGGTGTAGCGGTGAAATG
CGTAGAGATCAGGAGGAACATCGGTGGCGAAGGCGGCTTACTGGGCCTTTACTGA
CGCTGAGGCTCGAAAGCGTGGGGAGCAAAC

> Oligotype TGAAAAAAGATTTCGAACTAACGTC
TACATAGGGAGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGG
CTAAGCAAGTCAGATGTGAAATACACGGGCTCAACCCGTGAGCTGCATTTGAAACT
GTTTAGCTTGAGTGAAGTAGAGGCAGGCGGAATTCCCGGTGTAGCGGTGAAATGC
GTAGAGATCGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCTTTAACTGA
CGCTGAGGCACGAAAGCGTGGGTAGCAAAC

> Oligotype TTGGAGAAAATTTTGAACTTACGGC
TACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGTGTAGGCG
GGAAGGCAAGTCAGAAGTGAAAATTATGGGCTTAACCCATAACCTGCTTTTGAAAC
TGTTTTTCTTGAGTGAGGCAGAGGCAAGCGGAATTCCTAGTGTAGCGGTGAAATG
CGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTGCTGGGCCTTTACTGA
CGCTGAGGCTCGAAAGCGTGGGGAGCAAAC

> Oligotype TGAGTGAAAGTCTCGAACTTACGGC

TACGTAGGGAGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGTAGGCG
GGATGGCAAGTCAGATGTGAAAACTATGGGCTCAACCCATAGACTGCATTTGAAAC
TGCTGTTCTTGAGTGAGGTAGAGGTAAGCGGAATTCCTGGTGTAGCGGTGAAATG
CGTAGAGATCAGGAGGAACATCGGTGGCGAAGGCGGCTTACTGGGCCTTTACTGA
CGCTGAGGCTCGAAAGCGTGGGGAGCAAAC

> Oligotype TGAATTAAAATTTGGAACTAACGTA
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGG
CTTTGCAAGTCAGATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACT
GTAGAGCTTGAGTGAAGTAGAGGCAGGCGGAATTCCCCGTGTAGCGGTGAAATGC
GTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCTTTAACTGA
CGCTGAGGCACGAAAGCGTGGGTAGCAAAA

> Oligotype TGAATTAAAATTTGGAACTAACGGC
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGG
CTTTGCAAGTCAGATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACT
GTAGAGCTTGAGTGAAGTAGAGGCAGGCGGAATTCCCCGTGTAGCGGTGAAATGC
GTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCTTTAACTGA
CGCTGAGGCACGAAAGCGTGGGGAGCAAAC

PhyloSeq, *Ruminococceae Ruminococcus* sp. enriched in HIV-negative vs. CD4 <200

>Oligotype TACTCTGAACTAAACGTTC
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGG
CTTTGCAAGTCAGATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACT
GTAGAGCTTGAGTGAAGTAGAGGCAGGCGGAATTCCCCGTGTAGCGGTGAAATGC
GTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCTTTAACTGA
CGCTGAGGCACGAAAGCGTGGGTAGCAAAC

>Oligotype TAATGTCAACTAAACGTTC
TACATAGGGAGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGG
CTAAGCAAGTCAGATGTGAAATACACGGGCTCAACCCGTGAGCTGCATTTGAAACT
GTTTAGCTTGAGTGAAGTAGAGGCAGGCGGAATTCCCGGTGTAGCGGTGAAATGC
GTAGAGATCGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCTTTAACTGA
CGCTGAGGCACGAAAGCGTGGGTAGCAAAC

>Oligotype TACTCTGAACTAAACGTTA
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGG
CTTTGCAAGTCAGATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACT
GTAGAGCTTGAGTGAAGTAGAGGCAGGCGGAATTCCCCGTGTAGCGGTGAAATGC
GTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCTTTAACTGA
CGCTGAGGCACGAAAGCGTGGGTAGCAAAA

>Oligotype TACTCTGAACTAAACGTGC
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGG
CTTTGCAAGTCAGATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACT

GTAGAGCTTGAGTGAAGTAGAGGCAGGCGGAATTCCCCGTGTAGCGGTGAAATGC
GTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCTTTAACTGA
CGCTGAGGCACGAAAGCGTGGGGAGCAAAC

>Oligotype TATGATTAATCAAACATGC
TACGTAGGGAGCGAGCGTTGTCCGGAATTATTGGGTGTAAAGGGTGCGTAGGCG
GCTATGTAAGTCAGGCGTGTAATTCAGAGGCTTAACCTCTTGACGGCGCTTGAAAC
TGTGTAGCTTGAGTGGAGTAGAGGCAGATGGAATTTCCAGTGTAGCGGTGAAATG
CGTAGATATTGGAAGGAACATCGGTGGCGAAGGCGATCTGCTGGGCTCTAACTGA
CGCTGAGGCACGAAAGCATGGGGAGCAAAC

>Oligotype TACTCTGAATTAAACGTTC
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGG
CTTTGCAAGTCAGATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACT
GTAGAGCTTGAGTGAAGTAGAGGCAGGCGGAATTCCCCGTGTAGCGGTGAAATGC
GTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGTCTGCTGGGCTTTAACTGA
CGCTGAGGCACGAAAGCGTGGGTAGCAAAC

>Oligotype TATTATTAATTAAACATGC
TACGTAGGGAGCGAGCGTTGTCCGGAATTATTGGGTGTAAAGGGTGCGTAGGCG
GCTATGTAAGTCAGGCGTGTAATTCAGAGGCTTAACCTCTTGACTGCGCTTGAAAC
TGTGTAGCTTGAGTGGAGTAGAGGCAGATGGAATTTCCAGTGTAGCGGTGAAATG
CGTAGATATTGGAAGGAACATCGGTGGCGAAGGCGATCTGCTGGGCTTTAACTGA
CGCTGAGGCACGAAAGCATGGGGAGCAAAC

MaAsLin, *Enterobacteriaceae* enriched in CD4 <200

>Oligotype TTTCTTTTGTATGAC
TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCG
GTTTGTTAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATAC
TGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATG
CGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTG
ACGCTCAGGTGCGAAAGCGTGGGGAGCAAAC

>Oligotype TTTCTTTTGTATGAA
TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCG
GTTTGTTAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATAC
TGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATG
CGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTG
ACGCTCAGGTGCGAAAGCGTGGGGAGCAAAA

>Oligotype CTTCTTTTGTATGAC
CACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCG
GTTTGTTAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATAC
TGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATG

CGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTG
ACGCTCAGGTGCGAAAGCGTGGGGAGCAAAC

>Oligotype TTGTTTTTGTATGAC
TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCG
GTTGATTAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATAC
TGGTCAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATG
CGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTG
ACGCTCAGGTGCGAAAGCGTGGGGAGCAAAC

5.6 Statistical analysis (Chapter 2)

Descriptive measures were used to summarize the data. Continuous variables were summarized using median and IQR; categorical variables were summarized using frequency and percent (%). Spearman's rank correlations were used to examine bivariate associations between study variables. Fisher's exact and Chi-square tests were used to compare categorical variables between the study groups. Mann-Whitney test and Kruskal Wallis test (indicated by *p*-value in text) with Dunn's post hoc analyses (*p*-values in figures) were used for comparing continuous variables. PhyloSeq (v1.10.0) (McMurdie and Holmes, 2012) was used to calculate UniFrac distances between 16S samples and to perform principal coordinate analysis. MaAsLin (http://huttenhower.sph.harvard.edu/maaslin) was used for multivariate modeling by importing the relative bacterial abundance values and associated sample metadata. Minimum for feature relative abundance filtering was set to 1e-6, maximum false discovery rate at 0.05, minimum for feature prevalence filtering set to 0.1, and Benjamini-Hochberg FDR (BH-FDR) protocol was used for multiple comparison correction. Statistical significance of distance and dissimilarity metrics (beta-diversity) between groups was determined by PERMANOVA using the adonis function of QIIME. Differential abundance of bacterial taxa between experimental groups was determined

using the PhyloSeq DESeq2 extension using the Wald significance test and a parametric fit type (v.1.6.3) (Anders and Huber, 2010; McMurdie and Holmes, 2012) with multiple comparison correction using BH-FDR. No correction for multiple comparisons was performed unless otherwise stated. Statistical analyses and graphing were performed in R (Team, 2013) and Prism version 6.05 for Windows (GraphPad Software, La Jolla California USA, www.graphpad.com). All p-values were two-sided and $p < 0.05$ was considered significant.

5.7 Statistical analysis (Chapter 3)

Continuous variables were summarized using median and IQR; categorical variables were summarized using frequency and percent (%). Mann-Whitney test and Kruskal Wallis test (indicated by *p*-value in text) with Dunn's post hoc analyses (*p*-values in figures) were used for comparing continuous variables. Statistical significance of distance and dissimilarity metrics (beta-diversity) between groups was determined by PERMANOVA using the adonis function of vegan (v2.5.4) (Oksanen et al., 2013). Differential abundance of bacterial taxa between experimental groups was determined using the PhyloSeq DESeq2 (v.1.22.2) extension using the Wald significance test and a parametric fit type (Anders and Huber, 2010; McMurdie and Holmes, 2012) with multiple comparison correction using BH-FDR. No correction for multiple comparisons was performed unless otherwise stated. Statistical analyses and graphing were performed in R (Team, 2013). All p-values were two-sided and $p < 0.05$ was considered significant.

5.8 Isolation and culturing of primary *P. copri* isolates

Samples from these two cohorts were selected for *P. copri* isolation based on prior metagenomics 16S characterization, with preference given to samples associated with a high abundance of *P. copri*-associated 16S sequences.

We performed 0 to 10^{-7} dilutions in anaerobic conditions on Laked Blood Kanamycin Vancomycin (LKV) agar, which is selective for *Prevotella* and *Bacteroides*, and streaked unique colony morphologies for isolation on LKV. To distinguish between *Prevotella* and *Bacteroides*, we re-streaked isolated colonies onto Bacteroides Bile Esculin (BBE) agar and inspected for growth and pigmentation of the media. We chose BBE as a presumptive screening media because *P. copri* is known to hydrolyze esculin but does not grow in high bile concentrations. As such we hypothesized that the combination of no growth and production of black pigmentation from esculin hydrolysis would indicate that a colony was likely *P.copri.* All isolates were then picked from LKV into 3mL of Wilkins-Chalgren broth and incubated for 24-72 hours. Once a colony reached sufficiently dense growth, 1mL of the broth was pelleted for DNA extraction using a trizol and phenol-chloroform bead-beating based method and 1mL was mixed with 1X PBS plus 40% glycerol and frozen to create a working stock with a final concentration of 20% glycerol. A quick presumptive identification using 16S Sanger sequencing was performed.

We selected 21 samples likely to be enriched for *Prevotella*, 6 HIV+ and 5 HIV- samples from Boston and 5 HIV + and 5 HIV- samples from Uganda. Of the 21 samples, 19 grew on primary dilution plates. Of 284 colonies streaked for isolation, 278 grew and were re-streaked onto BBE and picked into broth for sequencing and storage. 254 grew in broth and of these, 109 were identified as *P. copri,* with a conservative

estimate of 27 unique, non-duplicate strains from 14 samples, 15 from the Boston cohort and 12 from the Ugandan cohort. The BBE readout test showed a sensitivity of 0.87 and specificity of 0.96.

5.9 Genomic DNA extraction and shotgun library preparation for bacterial isolates

Genomic DNA from culture isolates was extracted using a plate-based protocol including a bead beating process and combining phenol-chloroform isolation (Anahtar et al., 2016) with Qiagen QIAamp 96 DNA QIAcube HT kit procedures.

Shotgun library was prepared following a modified protocol of (Baym et al., 2015), using the Nextera DNA Library Preparation Kit (Illumina) and KAPA HiFi Library Amplification Kit (Kapa Biosystems). In brief, gDNA from each sample was standardized to concentration of 0.6ng/uL after quantification with SYBR Green I, followed by simultaneous fragmentation and sequencing adaptor incorporation by mixing 0.6ng gDNA (1uL) with 1.25uL TD buffer and 0.25uL TDE1 provided in the Nextera kit and incubating for 10min at 55°C. Tagmented DNA fragments were amplified in PCR using the KAPA high fidelity library amplification reagents, with Illumina adaptor sequences and sample barcodes incorporated in primers. PCR products were pooled, bead purified and sequenced on Illumina NextSeq with a 300-cycle kit.

5.10 Genome assembly and annotation for *P. copri* isolates

Sequences from Illumina NextSeq sequencing were 5' and 3' trimmed with a quality cutoff of 20 and a minimum read length of 45. Forward and reverse read pairs were assembled using SPAdes (Nurk et al., 2013). Genes were called using Prodigal (Hyatt et al., 2010). Assemblies were cut at 1.7 Mb based on reference genome and community genome sizes. Metaphlan2 (Segata et al., 2012) (and Bowtie2 (Langmead

and Salzberg, 2012)) were run on unassembled reads to assign taxonomy and determine contamination of genomes. Genes were clustered using CD-HIT (Li and Godzik, 2006) at 95% AA identity and annotated with eggNOG-Mapper (Huerta-Cepas et al., 2017).1 6s rRNA genes were extracted using barrnap (https://github.com/tseemann/barrnap) and checked for completeness by using BLAST (Altschul et al., 1990) on the beginning and end sequences. Marker genes were extracted using FetchMG (Mende et al., 2013) and FastTree (Price et al., 2010) and ETE3 (Huerta-Cepas et al., 2016) were used to align and draw trees. 16S rRNA gene trees were made using ClustalO (Sievers and Higgins, 2014) and RaxML (Stamatakis, 2014).

5.11 Comparative genomics of *P. copri*

P. copri genomes were clustered by binary Jaccard distance based on gene presence/absence and genomes with a pairwise distance of less than 0.40 were considered clones and collapsed. Dendrograms were constructed using ape (Paradis and Schliep, 2019) and the significance of clustering was determined using the mantel test against groups (mgroup) in the ecodist R package (Goslee and Urban, 2007). KEGG (Kanehisa and Goto, 2000) assignments of gene clusters in *P. copri* genomes were condensed into pathways and visualized using KEGGREST (Tenenbaum, 2018) and Pathview: (Luo and Brouwer, 2013). Significance of difference in pathway abundance was determined by Barnard's unconditional exact test (Barnard, 1947), which is considered a more powerful alternative to Fisher's exact test, implemented in the R package "Barnard".

References

Ahern, P.P., Faith, J.J., and Gordon, J.I. (2014). Mining the human gut microbiota for effector strains that shape the immune system. Immunity *40*, 815-823.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology *215*, 403-410.

Anahtar, M.N., Bowman, B.A., and Kwon, D.S. (2016). Efficient Nucleic Acid Extraction and 16S rRNA Gene Sequencing for Bacterial Community Characterization. Journal of visualized experiments : JoVE.

Anahtar, M.N., Gootenberg, D.B., Mitchell, C.M., and Kwon, D.S. (2018). Cervicovaginal Microbiota and Reproductive Health: The Virtue of Simplicity. Cell Host Microbe *23*, 159-168.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.

Andrews, C.A., and Koup, R.A. (1996). The immunopathology of HIV infection. The Journal of antimicrobial chemotherapy *37 Suppl B*, 13-25.

Angel, R. (2012). Total Nucleic Acid Extraction from Soil. Protocol Exchange.

Armstrong, A.J.S., Shaffer, M., Nusbacher, N.M., Griesmer, C., Fiorillo, S., Schneider, J.M., Preston Neff, C., Li, S.X., Fontenot, A.P., Campbell, T.*, et al.* (2018). An exploration of Prevotella-rich microbiomes in HIV and men who have sex with men. Microbiome *6*, 198.

Atarashi, K., Tanoue, T., Oshima, K., Suda, W., Nagano, Y., Nishikawa, H., Fukuda, S., Saito, T., Narushima, S., Hase, K.*, et al.* (2013). Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. Nature *500*, 232-236.

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

Ayeni, F.A., Biagi, E., Rampelli, S., Fiori, J., Soverini, M., Audu, H.J., Cristino, S., Caporali, L., Schnorr, S.L., Carelli, V.*, et al.* (2018). Infant and Adult Gut Microbiome and Metabolome in Rural Bassa and Urban Settlers from Nigeria. Cell reports *23*, 3056-3067.

Backhed, F., Fraser, C.M., Ringel, Y., Sanders, M.E., Sartor, R.B., Sherman, P.M., Versalovic, J., Young, V., and Finlay, B.B. (2012). Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. Cell Host Microbe *12*, 611-622.

Bailey, M.T., Dowd, S.E., Parry, N.M., Galley, J.D., Schauer, D.B., and Lyte, M. (2010). Stressor exposure disrupts commensal microbial populations in the intestines and leads to increased colonization by Citrobacter rodentium. Infection and immunity *78*, 1509-1519.

Bajaj, J.S., Hylemon, P.B., Ridlon, J.M., Heuman, D.M., Daita, K., White, M.B., Monteith, P., Noble, N.A., Sikaroodi, M., and Gillevet, P.M. (2012). Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. American journal of physiology Gastrointestinal and liver physiology *303*, G675-685.

Barnard, G.A. (1947). Significance tests for 2 X 2 tables. Biometrika *34*, 123-138.

Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M., and Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. PLoS One *10*, e0128036.

Bender, J.M., Li, F., Martelly, S., Byrt, E., Rouzier, V., Leo, M., Tobin, N., Pannaraj, P.S., Adisetiyo, H., Rollie, A.*, et al.* (2016). Maternal HIV infection influences the microbiome of HIV-uninfected infants. Sci Transl Med *8*, 349ra100.

Bennett, B.J., Davis, R.C., Civelek, M., Orozco, L., Wu, J., Qi, H., Pan, C., Packard, R.R., Eskin, E., Yan, M.*, et al.* (2015). Genetic Architecture of Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. PLoS genetics *11*, e1005711.

Bereswill, S., Fischer, A., Plickert, R., Haag, L.M., Otto, B., Kuhl, A.A., Dasti, J.I., Zautner, A.E., Munoz, M., Loddenkemper, C.*, et al.* (2011). Novel murine infection models provide deep insights into the "menage a trois" of Campylobacter jejuni, microbiota and host innate immunity. PLoS One *6*, e20953.

Bhavani, A.L., and Nisha, J. (2010). Dextran - The polysaccharide with versatile uses, Vol 1.

Biddle, A., Stewart, L., Blanchard, J., and Leschine, S. (2013). Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities. Diversity *5*, 627.

Blaser, M.J. (2016). Antibiotic use and its consequences for the normal microbiome. Science *352*, 544-545.

Blaser, M.J. (2018). The Past and Future Biology of the Human Microbiome in an Age of Extinctions. Cell *172*, 1173-1177.

Blekhman, R., Tang, K., Archie, E.A., Barreiro, L.B., Johnson, Z.P., Wilson, M.E., Kohn, J., Yuan, M.L., Gesquiere, L., Grieneisen, L.E.*, et al.* (2016). Common methods for fecal sample storage in field studies yield consistent signatures of individual identity in microbiome sequencing data. Scientific reports *6*, 31519.

Bokulich, N.A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., A, D.L., Wu, F., Perez-Perez, G.I., Chen, Y.*, et al.* (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci Transl Med *8*, 343ra382.

Brenchley, J.M. (2013). Mucosal immunity in human and simian immunodeficiency lentivirus infections. Mucosal Immunol *6*, 657-665.

Brenchley, J.M., and Douek, D.C. (2012). Microbial translocation across the GI tract. Annu Rev Immunol *30*, 149-173.

Brenchley, J.M., Price, D.a., Schacker, T.W., Asher, T.E., Silvestri, G., Rao, S., Kazzaz, Z., Bornstein, E., Lambotte, O., Altmann, D.*, et al.* (2006). Microbial translocation is a cause of systemic immune activation in chronic HIV infection. Nature medicine *12*, 1365-1371.

Brenchley, J.M., Schacker, T.W., Ruff, L.E., Price, D.A., Taylor, J.H., Beilman, G.J., Nguyen, P.L., Khoruts, A., Larson, M., Haase, A.T.*, et al.* (2004). CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. The Journal of experimental medicine *200*, 749-759.

Brown, J.M., and Hazen, S.L. (2015). The Gut Microbial Endocrine Organ: Bacterially Derived Signals Driving Cardiometabolic Diseases. Annual Review of Medicine *66*, 343-359.

Cadwell, K., Patel, K.K., Maloney, N.S., Liu, T.C., Ng, A.C., Storer, C.E., Head, R.D., Xavier, R., Stappenbeck, T.S., and Virgin, H.W. (2010). Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. Cell *141*, 1135-1145.

Cailhier, J.F., Partolina, M., Vuthoori, S., Wu, S., Ko, K., Watson, S., Savill, J., Hughes, J., and Lang, R.A. (2005). Conditional macrophage ablation demonstrates that resident macrophages initiate acute peritoneal inflammation. Journal of immunology *174*, 2336-2342.

Callahan, B.J., McMurdie, P.J., and Holmes, S.P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME journal *11*, 2639-2643.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods *13*, 581-583.

Candon, S., Perez-Arroyo, A., Marquet, C., Valette, F., Foray, A.P., Pelletier, B., Milani, C., Ventura, M., Bach, J.F., and Chatenoud, L. (2015). Antibiotics in early life alter the gut microbiome and increase disease incidence in a spontaneous mouse model of autoimmune insulin-dependent diabetes. PLoS One *10*, e0125448.

Cani, P.D., Amar, J., Iglesias, M.A., Poggi, M., Knauf, C., Bastelica, D., Neyrinck, A.M., Fava, F., Tuohy, K.M., Chabo, C.*, et al.* (2007). Metabolic endotoxemia initiates obesity and insulin resistance. Diabetes *56*, 1761-1772.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., and Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics *26*, 266-267.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I.*, et al.* (2010b). QIIME allows analysis of high-throughput community sequencing data. Nature Methods *7*, 335-336.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences of the United States of America *108*, 4516-4522.

Cassol, E., Malfeld, S., Mahasha, P., van der Merwe, S., Cassol, S., Seebregts, C., Alfano, M., Poli, G., and Rossouw, T. (2010). Persistent microbial translocation and immune activation in HIV-1-infected South Africans receiving combination antiretroviral therapy. J Infect Dis *202*, 723-733.

Centers for Disease Control and Prevention (2008). Centers for Disease Control and Prevention: HIV Surveillance Report.

Centers for Disease Control and Prevention (2019). Estimated HIV incidence and prevalence in the United States, 2010–2016. HIV Surveillance Supplemental Report *24*.

Chang, J.Y., Antonopoulos, D.A., Kalra, A., Tonelli, A., Khalife, W.T., Schmidt, T.M., and Young, V.B. (2008). Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea. J Infect Dis *197*, 435-438.

Chen, W., Liu, F., Ling, Z., Tong, X., and Xiang, C. (2012). Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PloS one *7*, e39743.

Cho, I., and Blaser, M.J. (2012). The human microbiome: at the interface of health and disease. Nature reviews Genetics *13*, 260-270.

Chow, F.C., Regan, S., Feske, S., Meigs, J.B., Grinspoon, S.K., and Triant, V.A. (2012). Comparison of ischemic stroke incidence in HIV-infected and non-HIV-infected patients in a US health care system. Journal of acquired immune deficiency syndromes (1999) *60*, 351-358.

Clayton, F., Snow, G., Reka, S., and Kotler, D.P. (1997). Selective depletion of rectal lamina propria rather than lymphoid aggregate CD4 lymphocytes in HIV infection. Clinical and experimental immunology *107*, 288-292.

Collaboration of Observational, H.I.V.E.R.E.i.E., Lewden, C., Bouteloup, V., De Wit, S., Sabin, C., Mocroft, A., Wasmuth, J.C., van Sighem, A., Kirk, O., Obel, N.*, et al.* (2012). All-cause mortality in treated HIV-infected adults with CD4 >/=500/mm3

compared with the general population: evidence from a large European observational cohort collaboration. International journal of epidemiology *41*, 433-445.

Cox, L.M., Yamanishi, S., Sohn, J., Alekseyenko, A.V., Leung, J.M., Cho, I., Kim, S.G., Li, H., Gao, Z., Mahana, D.*, et al.* (2014). Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. Cell *158*, 705-721.

Crost, E.H., Tailford, L.E., Le Gall, G., Fons, M., Henrissat, B., and Juge, N. (2013). Utilisation of mucin glycans by the human gut symbiont Ruminococcus gnavus is strain-dependent. PLoS One *8*, e76341.

Cunningham-Rundles, S., Ahrne, S., Johann-Liang, R., Abuav, R., Dunn-Navarra, A.M., Grassey, C., Bengmark, S., and Cervia, J.S. (2011). Effect of probiotic bacteria on microbial host defense, growth, and immune function in human immunodeficiency virus type-1 infection. Nutrients *3*, 1042-1070.

d'Ettorre, G., Ceccarelli, G., Giustini, N., Serafino, S., Calantone, N., De Girolamo, G., Bianchi, L., Bellelli, V., Ascoli-Bartoli, T., Marcellini, S.*, et al.* (2015). Probiotics Reduce Inflammation in Antiretroviral Treated, HIV-Infected Individuals: Results of the "Probio-HIV" Clinical Trial. PLoS One *10*, e0137200.

David, L.a., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A.*, et al.* (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559-563.

de Bentzmann, S., and Plesiat, P. (2011). The Pseudomonas aeruginosa opportunistic pathogen and human infections. Environmental microbiology *13*, 1655-1665.

de Bonnecaze, G., Chaput, B., Dupret-Bories, A., Vergez, S., and Serrano, E. (2018). Functional outcome after long-term low-dose trimethoprim/sulfamethoxazole in chronic rhinosinusitis with purulence: a prospective study. The Journal of laryngology and otology *132*, 600-604.

De Filippis, F., Pasolli, E., Tett, A., Tarallo, S., Naccarati, A., De Angelis, M., Neviani, E., Cocolin, L., Gobbetti, M., Segata, N.*, et al.* (2019). Distinct Genetic and Functional Traits of Human Intestinal Prevotella copri Strains Are Associated with Different Habitual Diets. Cell Host Microbe *25*, 444-453 e443.

De Filippis, F., Pellegrini, N., Vannini, L., Jeffery, I.B., La Storia, A., Laghi, L., Serrazanetti, D.I., Di Cagno, R., Ferrocino, I., Lazzi, C.*, et al.* (2016). High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. Gut *65*, 1812-1821.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proceedings of the National Academy of Sciences of the United States of America *107*, 14691-14696.

Delwart, E. (2013). A roadmap to the human virome. PLoS pathogens *9*, e1003146.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology *72*, 5069-5072.

Dillon, S.M., Kibbie, J., Lee, E.J., Guo, K., Santiago, M.L., Austin, G.L., Gianella, S., Landay, A.L., Donovan, A.M., Frank, D.N.*, et al.* (2017). Low abundance of colonic butyrate-producing bacteria in HIV infection is associated with microbial translocation and immune activation. Aids *31*, 511-521.

Dillon, S.M., Lee, E.J., Kotter, C.V., Austin, G.L., Dong, Z., Hecht, D.K., Gianella, S., Siewe, B., Smith, D.M., Landay, A.L.*, et al.* (2014). An altered intestinal mucosal microbiome in HIV-1 infection is associated with mucosal and systemic immune activation and endotoxemia. Mucosal immunology *7*, 983-994.

Dillon, S.M., Lee, E.J., Kotter, C.V., Austin, G.L., Gianella, S., Siewe, B., Smith, D.M., Landay, A.L., McManus, M.C., Robertson, C.E.*, et al.* (2016). Gut dendritic cell activation links an altered colonic microbiome to mucosal and systemic T-cell activation in untreated HIV-1 infection. Mucosal Immunol *9*, 24-37.

Dinh, D.M., Volpe, G.E., Duffalo, C., Bhalchandra, S., Tai, A.K., Kane, A.V., Wanke, C.a., and Ward, H.D. (2015). Intestinal microbiota, microbial translocation, and systemic inflammation in chronic HIV infection. The Journal of infectious diseases *211*, 19-27.

Donaldson, G.P., Lee, S.M., and Mazmanian, S.K. (2016). Gut biogeography of the bacterial microbiota. Nat Rev Microbiol *14*, 20-32.

Douek, D.C., Roederer, M., and Koup, R.A. (2009). Emerging concepts in the immunopathogenesis of AIDS. Annual review of medicine *60*, 471-484.

Dubourg, G., Lagier, J.C., Hue, S., Surenaud, M., Bachar, D., Robert, C., Michelle, C., Ravaux, I., Mokhtari, S., Million, M.*, et al.* (2016). Gut microbiota associated with HIV infection is significantly enriched in bacteria tolerant to oxygen. BMJ open gastroenterology *3*, e000080.

Duprez, D.A., Neuhaus, J., Kuller, L.H., Tracy, R., Belloso, W., De Wit, S., Drummond, F., Lane, H.C., Ledergerber, B., Lundgren, J.*, et al.* (2012). Inflammation, coagulation and cardiovascular disease in HIV-infected individuals. PLoS ONE *7*.

Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nature communications *8*, 1784.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. (2005). Diversity of the human intestinal microbial flora. Science *308*, 1635-1638.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460-2461.

Egger, M., May, M., Chene, G., Phillips, A.N., Ledergerber, B., Dabis, F., Costagliola, D., D'Arminio Monforte, A., de Wolf, F., Reiss, P.*, et al.* (2002). Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. Lancet *360*, 119-129.

Ellis, C.L., Ma, Z.M., Mann, S.K., Li, C.S., Wu, J., Knight, T.H., Yotter, T., Hayes, T.L., Maniar, A.H., Troia-Cancio, P.V.*, et al.* (2011). Molecular characterization of stool microbiota in HIV-infected subjects by panbacterial and order-level 16S ribosomal DNA (rDNA) quantification and correlations with immune activation. J Acquir Immune Defic Syndr *57*, 363-370.

Eren, A.M., Zozaya, M., Taylor, C.M., Dowd, S.E., Martin, D.H., and Ferris, M.J. (2011). Exploring the diversity of Gardnerella vaginalis in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. PLoS One *6*, e26732.

Faith, D.P., and Baker, A.M. (2006). Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. Evolutionary Bioinformatics Online *2*, 121-128.

Faith, J.J., Colombel, J.F., and Gordon, J.I. (2015). Identifying strains that contribute to complex diseases through the study of microbial inheritance. Proc Natl Acad Sci U S A *112*, 633-640.

Federico, A., Tuccillo, C., Grossi, E., Abbiati, R., Garbagna, N., Romano, M., Tiso, A., Blanco Cdel, V., and Loguercio, C. (2009). The effect of a new symbiotic formulation on plasma levels and peripheral blood mononuclear cell expression of some pro-inflammatory cytokines in patients with ulcerative colitis: a pilot study. European review for medical and pharmacological sciences *13*, 285-293.

Ferreira, R.B., Gill, N., Willing, B.P., Antunes, L.C., Russell, S.L., Croxen, M.A., and Finlay, B.B. (2011). The intestinal microbiota plays a role in Salmonella-induced colitis independent of pathogen colonization. PLoS One *6*, e20338.

Fettig, J., Swaminathan, M., Murrill, C.S., and Kaplan, J.E. (2014). Global epidemiology of HIV. Infectious disease clinics of North America *28*, 323-337.

Fisher, E.A., Feig, J.E., Hewing, B., Hazen, S.L., and Smith, J.D. (2012). High-density lipoprotein function, dysfunction, and reverse cholesterol transport. Arteriosclerosis, thrombosis, and vascular biology *32*, 2813-2820.

Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., and White, B.A. (2008). Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol *6*, 121-131.

Franke, T., and Deppenmeier, U. (2018). Physiology and central carbon metabolism of the gut bacterium Prevotella copri. Molecular microbiology *109*, 528-540.

Franzosa, E.A., McIver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N.*, et al.* (2018). Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods *15*, 962-968.

Freiberg, M.S., Chang, C.-C.H., Kuller, L.H., Skanderson, M., Lowy, E., Kraemer, K.L., Butt, A.a., Bidwell Goetz, M., Leaf, D., Oursler, K.A.*, et al.* (2013). HIV infection and the risk of acute myocardial infarction. JAMA internal medicine *173*, 614-622.

Friedrich, M.J. (2013). Genomes of microbes inhabiting the body offer clues to human health and disease. JAMA *309*, 1447-1449.

Fulcher, J.A., Hussain, S.K., Cook, R., Li, F., Tobin, N.H., Ragsdale, A., Shoptaw, S., Gorbach, P.M., and Aldrovandi, G.M. (2018). Effects of Substance Use and Sex Practices on the Intestinal Microbiome During HIV-1 Infection. J Infect Dis *218*, 1560-1570.

Furusawa, Y., Obata, Y., Fukuda, S., Endo, T.A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T.*, et al.* (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. Nature *504*, 446-450.

Gandhi, R.T., Spritzler, J., Chan, E., Asmuth, D.M., Rodriguez, B., Merigan, T.C., Hirsch, M.S., Shafer, R.W., Robbins, G.K., and Pollard, R.B. (2006). Effect of baseline- and treatment-related factors on immunologic recovery after initiation of antiretroviral therapy in HIV-1-positive subjects: results from ACTG 384. J Acquir Immune Defic Syndr *42*, 426-434.

Gauffin Cano, P., Santacruz, A., Moya, A., and Sanz, Y. (2012). Bacteroides uniformis CECT 7771 ameliorates metabolic and immunological dysfunction in mice with high-fat-diet induced obesity. PLoS One *7*, e41079.

Gevers, D., Kugathasan, S., Denson, L.A., Vazquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M.*, et al.* (2014). The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe *15*, 382-392.

Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., and Nelson, K.E. (2006). Metagenomic analysis of the human distal gut microbiome. Science *312*, 1355-1359.

Global Burden of Disease Collaborative Network (2018). Global Burden of Disease Study 2017 (GBD 2017) Results.

Goicoechea, M., Smith, D.M., Liu, L., May, S., Tenorio, A.R., Ignacio, C.C., Landay, A., and Haubrich, R. (2006). Determinants of CD4+ T cell recovery during suppressive antiretroviral therapy: association of immune activation, T cell maturation markers, and cellular HIV-1 DNA. The Journal of infectious diseases *194*, 29-37.

Gomez, A., Petrzelkova, K.J., Burns, M.B., Yeoman, C.J., Amato, K.R., Vlckova, K., Modry, D., Todd, A., Jost Robinson, C.A., Remis, M.J.*, et al.* (2016). Gut Microbiome of Coexisting BaAka Pygmies and Bantu Reflects Gradients of Traditional Subsistence Patterns. Cell reports *14*, 2142-2153.

González-Hernández, L.A., Jave-Suarez, L.F., Fafutis-Morris, M., Montes-Salcedo, K.E., Valle-Gutierrez, L.G., Campos-Loza, A.E., Enciso-Gómez, L.F., and Andrade-Villanueva, J.F. (2012). Synbiotic therapy decreases microbial translocation and inflammation and improves immunological status in HIV-infected patients: a double-blind randomized controlled pilot trial. Nutrition journal *11*, 90.

Gootenberg, D.B., Paer, J.M., Luevano, J.M., and Kwon, D.S. (2017). HIV-associated changes in the enteric microbial community: potential role in loss of homeostasis and development of systemic inflammation. Current opinion in infectious diseases *30*, 31-43.

Gori, A., Tincati, C., Rizzardini, G., Torti, C., Quirino, T., Haarman, M., Ben Amor, K., van Schaik, J., Vriesema, A., Knol, J.*, et al.* (2008). Early impairment of gut function and gut flora supporting a role for alteration of gastrointestinal mucosa in human immunodeficiency virus pathogenesis. Journal of clinical microbiology *46*, 757-758.

Gorvitovskaia, A., Holmes, S.P., and Huse, S.M. (2016). Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. Microbiome *4*, 15.

Goslee, S.C., and Urban, D.L. (2007). The ecodist Package for Dissimilarity-based Analysis of Ecological Data. 2007 *22*, 19.

Gregory, J.C., Buffa, J.A., Org, E., Wang, Z., Levison, B.S., Zhu, W., Wagner, M.A., Bennett, B.J., Li, L., DiDonato, J.A.*, et al.* (2015). Transmission of atherosclerosis susceptibility with gut microbial transplantation. The Journal of biological chemistry *290*, 5647-5660.

Greiner, T., and Backhed, F. (2011). Effects of the gut microbiota on obesity and glucose homeostasis. Trends in endocrinology and metabolism: TEM *22*, 117-123.

Grunfeld, C., Delaney, J.A., Wanke, C., Currier, J.S., Scherzer, R., Biggs, M.L., Tien, P.C., Shlipak, M.G., Sidney, S., Polak, J.F.*, et al.* (2009). Preclinical

atherosclerosis due to HIV infection: carotid intima-medial thickness measurements from the FRAM study. Aids *23*, 1841-1849.

Guaraldi, G., Orlando, G., Zona, S., Menozzi, M., Carli, F., Garlassi, E., Berti, A., Rossi, E., Roverato, A., and Palella, F. (2011). Premature age-related comorbidities among HIV-infected persons compared with the general population. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America *53*, 1120-1126.

Guentzel, M.N. (1996). Escherichia, Klebsiella, Enterobacter, Serratia, Citrobacter, and Proteus. In Medical Microbiology, S. Baron, ed. (Galveston (TX)).

Haaland, R.E., Fountain, J., Hu, Y., Holder, A., Dinh, C., Hall, L., Pescatore, N.A., Heeke, S., Hart, C.E., Xu, J.*, et al.* (2018). Repeated rectal application of a hyperosmolar lubricant is associated with microbiota shifts but does not affect PrEP drug concentrations: results from a randomized trial in men who have sex with men. Journal of the International AIDS Society *21*, e25199.

Hale, T.L., and Keusch, G.T. (1996). Shigella.

Haller, D., Bode, C., Hammes, W.P., Pfeifer, A.M., Schiffrin, E.J., and Blum, S. (2000). Non-pathogenic bacteria elicit a differential cytokine response by intestinal epithelial cell/leucocyte co-cultures. Gut *47*, 79-87.

Handley, S.A., Desai, C., Zhao, G., Droit, L., Monaco, C.L., Schroeder, A.C., Nkolola, J.P., Norman, M.E., Miller, A.D., Wang, D.*, et al.* (2016). SIV Infection-Mediated Changes in Gastrointestinal Bacterial Microbiome and Virome Are Associated with Immunodeficiency and Prevented by Vaccination. Cell Host Microbe *19*, 323-335.

Handley, S.A., Thackray, L.B., Zhao, G., Presti, R., Miller, A.D., Droit, L., Abbink, P., Maxfield, L.F., Kambal, A., Duan, E.*, et al.* (2012). Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. Cell *151*, 253-266.

Hara, N., Alkanani, A.K., Ir, D., Robertson, C.E., Wagner, B.D., Frank, D.N., and Zipris, D. (2012). Prevention of virus-induced type 1 diabetes with antibiotic therapy. Journal of immunology *189*, 3805-3814.

Harris, K., Kassis, A., Major, G., and Chou, C.J. (2012). Is the gut microbiota a new factor contributing to obesity and its metabolic disorders? Journal of obesity *2012*, 879151.

Hartig, E., Geng, H., Hartmann, A., Hubacek, A., Munch, R., Ye, R.W., Jahn, D., and Nakano, M.M. (2004). Bacillus subtilis ResD induces expression of the potential regulatory genes yclJK upon oxygen limitation. Journal of bacteriology *186*, 6477-6484.

Hayashi, H., Shibata, K., Sakamoto, M., Tomita, S., and Benno, Y. (2007). Prevotella copri sp. nov. and Prevotella stercorea sp. nov., isolated from human faeces. International journal of systematic and evolutionary microbiology *57*, 941-946.

Heise, C., Miller, C.J., Lackner, A., and Dandekar, S. (1994). Primary acute simian immunodeficiency virus infection of intestinal lymphoid tissue is associated with gastrointestinal dysfunction. The Journal of infectious diseases *169*, 1116-1120.

Hirschhorn, L.R., Kaaya, S.F., Garrity, P.S., Chopyak, E., and Fawzi, M.C.S. (2012). Cancer and the 'other' noncommunicable chronic diseases in older people living with HIV/AIDS in resource-limited settings: a challenge to success. AIDS (London, England) *26 Suppl 1*, S65-75.

HIV/AIDS, J.U.N.P.o. (2013). Global report: UNAIDS report on the global AIDS epidemic 2013. (WHO Library Cataloguing-in-Publication Data).

Hoel, H., Hove-Skovsgaard, M., Hov, J.R., Gaardbo, J.C., Holm, K., Kummen, M., Rudi, K., Nwosu, F., Valeur, J., Gelpi, M.*, et al.* (2018). Impact of HIV and Type 2 diabetes on Gut Microbiota Diversity, Tryptophan Catabolism and Endothelial Dysfunction. Scientific reports *8*, 6725.

Honda, K., and Littman, D.R. (2012). The microbiome in infectious disease and inflammation. Annual review of immunology *30*, 759-795.

Hontelez, J.A.C., de Vlas, S.J., Baltussen, R., Newell, M.-L., Bakker, R., Tanser, F., Lurie, M., and Bärnighausen, T. (2012). The impact of antiretroviral treatment on the age composition of the HIV epidemic in sub-Saharan Africa. AIDS (London, England) *26 Suppl 1*, S19-30.

Hooper, L.V., Littman, D.R., and Macpherson, A.J. (2012). Interactions between the microbiota and the immune system. Science *336*, 1268-1273.

Hooper, L.V., and Macpherson, A.J. (2010). Immune adaptations that maintain homeostasis with the intestinal microbiota. Nature reviews Immunology *10*, 159-169.

Hsiao, A., Ahmed, A.M., Subramanian, S., Griffin, N.W., Drewry, L.L., Petri, W.A., Jr., Haque, R., Ahmed, T., and Gordon, J.I. (2014). Members of the human gut microbiota involved in recovery from Vibrio cholerae infection. Nature *515*, 423-426.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Molecular biology and evolution *34*, 2115-2122.

Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Molecular biology and evolution *33*, 1635-1638.

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

Hummelen, R., Hemsworth, J., Changalucha, J., Butamanya, N.L., Hekmat, S., Habbema, J.D., and Reid, G. (2011). Effect of micronutrient and probiotic fortified yogurt on immune-function of anti-retroviral therapy naive HIV patients. Nutrients *3*, 897-909.

Hunt, P.W. (2012). HIV and inflammation: mechanisms and consequences. Current HIV/AIDS reports *9*, 139-147.

Hunt, Peter W., Brenchley, J., Sinclair, E., McCune, Joseph M., Roland, M., Page‐Shafer, K., Hsue, P., Emu, B., Krone, M., Lampiris, H.*, et al.* (2008). Relationship between T Cell Activation and CD4 + T Cell Count in HIV‐Seropositive Individuals with Undetectable Plasma HIV RNA Levels in the Absence of Therapy. The Journal of Infectious Diseases *197*, 126-133.

Hunt, P.W., Cao, H.L., Muzoora, C., Ssewanyana, I., Bennett, J., Emenyonu, N., Kembabazi, A., Neilands, T.B., Bangsberg, D.R., Deeks, S.G.*, et al.* (2011). Impact of CD8+ T-cell activation on CD4+ T-cell recovery and mortality in HIV-infected Ugandans initiating antiretroviral therapy. AIDS (London, England) *25*, 2123-2131.

Hunt, P.W., Martin, J.N., Sinclair, E., Bredt, B., Hagos, E., Lampiris, H., and Deeks, S.G. (2003). T cell activation is associated with lower CD4+ T cell gains in human immunodeficiency virus-infected patients with sustained viral suppression during antiretroviral therapy. The Journal of infectious diseases *187*, 1534-1543.

Hunt, P.W., Sinclair, E., Rodriguez, B., Shive, C., Clagett, B., Funderburg, N., Robinson, J., Huang, Y., Epling, L., Martin, J.N.*, et al.* (2014). Gut epithelial barrier dysfunction and innate immune activation predict mortality in treated HIV infection. The Journal of infectious diseases *210*, 1228-1238.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics *11*, 119.

Igarashi, T., Yamamoto, A., and Goto, N. (1998). Detection of dextranase-producing gram-negative oral bacteria. Oral microbiology and immunology *13*, 382-386.

Irvine, S.L., Hummelen, R., Hekmat, S., Looman, C.W., Habbema, J.D., and Reid, G. (2010). Probiotic yogurt consumption is associated with an increase of CD4 count among people living with HIV/AIDS. Journal of clinical gastroenterology *44*, e201-205.

Islam, K.B., Fukiya, S., Hagio, M., Fujii, N., Ishizuka, S., Ooka, T., Ogura, Y., Hayashi, T., and Yokota, A. (2011). Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. Gastroenterology *141*, 1773-1781.

Ivanov, II, Atarashi, K., Manel, N., Brodie, E.L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K.C., Santee, C.A., Lynch, S.V.*, et al.* (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell *139*, 485-498.

Jenabian, M.A., El-Far, M., Vyboh, K., Kema, I., Costiniuk, C.T., Thomas, R., Baril, J.G., LeBlanc, R., Kanagaratham, C., Radzioch, D.*, et al.* (2015). Immunosuppressive Tryptophan Catabolism and Gut Mucosal Dysfunction Following Early HIV Infection. J Infect Dis *212*, 355-366.

Jiang, W., Lederman, M.M., Hunt, P., Sieg, S.F., Haley, K., Rodriguez, B., Landay, A., Martin, J., Sinclair, E., Asher, A.I.*, et al.* (2009). Plasma levels of bacterial DNA correlate with immune activation and the magnitude of immune restoration in persons with antiretroviral-treated HIV infection. J Infect Dis *199*, 1177-1185.

Justice, A.C., and Braithwaite, R.S. (2012). Lessons learned from the first wave of aging with HIV. AIDS (London, England) *26 Suppl 1*, S11-18.

Kaakoush, N.O. (2015). Insights into the Role of Erysipelotrichaceae in the Human Host. Frontiers in cellular and infection microbiology *5*, 84.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

Kang, S., Denman, S.E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., and McSweeney, C.S. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. Inflammatory bowel diseases *16*, 2034-2042.

Katz, I.T., and Maughan-Brown, B. (2017). Improved life expectancy of people living with HIV: who is left behind? The lancet HIV *4*, e324-e326.

Kehrmann, J., Menzel, J., Saeedghalati, M., Obeid, R., Schulze, C., Holzendorf, V., Farahpour, F., Reinsch, N., Klein-Hitpass, L., Streeck, H.*, et al.* (2019). Gut Microbiota in Human Immunodeficiency Virus-Infected Individuals Linked to Coronary Heart Disease. J Infect Dis *219*, 497-508.

Kelley, C.F., Kraft, C.S., de Man, T.J., Duphare, C., Lee, H.W., Yang, J., Easley, K.A., Tharp, G.K., Mulligan, M.J., Sullivan, P.S.*, et al.* (2017). The rectal mucosa and condomless receptive anal intercourse in HIV-negative MSM: implications for HIV transmission and prevention. Mucosal Immunol *10*, 996-1007.

Kenyon, C., and Osbak, K. (2014). Certain attributes of the sexual ecosystem of high-risk MSM have resulted in an altered microbiome with an enhanced propensity to generate and transmit antibiotic resistance. Medical hypotheses *83*, 196-202.

Khanna, S., Montassier, E., Schmidt, B., Patel, R., Knights, D., Pardi, D.S., and Kashyap, P.C. (2016). Gut microbiome predictors of treatment response and recurrence in primary Clostridium difficile infection. Alimentary pharmacology & therapeutics *44*, 715-727.

Khovidhunkit, W., Kim, M.-S., Memon, R.A., Shigenaga, J.K., Moser, A.H., Feingold, K.R., and Grunfeld, C. (2004). Effects of infection and inflammation on lipid and lipoprotein metabolism: mechanisms and consequences to the host. Journal of lipid research *45*, 1169-1196.

Kim, H.J., Li, H., Collins, J.J., and Ingber, D.E. (2016). Contributions of microbiome and mechanical deformation to intestinal bacterial overgrowth and inflammation in a human gut-on-a-chip. Proc Natl Acad Sci U S A *113*, E7-15.

Kitchens, R.L. (2000). Role of CD14 in cellular recognition of bacterial lipopolysaccharides. Chemical immunology *74*, 61-82.

Klatt, N.R., Villinger, F., Bostik, P., Gordon, S.N., Pereira, L., Engram, J.C., Mayne, A., Dunham, R.M., Lawson, B., Ratcliffe, S.J.*, et al.* (2008). Availability of activated CD4+ T cells dictates the level of viremia in naturally SIV-infected sooty mangabeys. J Clin Invest *118*, 2039-2049.

Knapp, S., Wieland, C.W., Florquin, S., Pantophlet, R., Dijkshoorn, L., Tshimbalanga, N., Akira, S., and van der Poll, T. (2006). Differential roles of CD14 and toll-like receptors 4 and 2 in murine Acinetobacter pneumonia. American journal of respiratory and critical care medicine *173*, 122-129.

Koeth, R.A., Levison, B.S., Culley, M.K., Buffa, J.A., Wang, Z., Gregory, J.C., Org, E., Wu, Y., Li, L., Smith, J.D.*, et al.* (2014). γ -butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO. Cell Metabolism *20*, 799-812.

Koeth, R.a., Wang, Z., Levison, B.S., Buffa, J.a., Org, E., Sheehy, B.T., Britt, E.B., Fu, X., Wu, Y., Li, L.*, et al.* (2013). Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nature medicine *19*, 576-585.

Kofteridis, D.P., Maraki, S., Mixaki, I., Mantadakis, E., and Samonis, G. (2004). Impact of prolonged treatment with trimethoprim-sulfamethoxazole on the human gut flora. Scandinavian journal of infectious diseases *36*, 771-772.

Koler, M., Frank, V., Amartely, H., Friedler, A., and Vaknin, A. (2016). Dynamic Clustering of the Bacterial Sensory Kinase BaeS. PLoS One *11*, e0150349.

Koropatkin, N.M., Cameron, E.A., and Martens, E.C. (2012). How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol *10*, 323-335.

Kostic, A.D., Chun, E., Robertson, L., Glickman, J.N., Gallini, C.A., Michaud, M., Clancy, T.E., Chung, D.C., Lochhead, P., Hold, G.L.*, et al.* (2013). Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe *14*, 207-215.

Kostic, A.D., Gevers, D., Siljander, H., Vatanen, T., Hyotylainen, T., Hamalainen, A.M., Peet, A., Tillmann, V., Poho, P., Mattila, I.*, et al.* (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe *17*, 260-273.

Kostic, A.D., Xavier, R.J., and Gevers, D. (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology *146*, 1489-1499.

Kovatcheva-Datchary, P., Nilsson, A., Akrami, R., Lee, Y.S., De Vadder, F., Arora, T., Hallen, A., Martens, E., Bjorck, I., and Backhed, F. (2015). Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. Cell Metab *22*, 971-982.

Kunisawa, J., and Kiyono, H. (2011). Peaceful mutualism in the gut: revealing key commensal bacteria for the creation and maintenance of immunological homeostasis. Cell Host Microbe *9*, 83-84.

Kunjathoor, V.V., Febbraio, M., Podrez, E.A., Moore, K.J., Andersson, L., Koehn, S., Rhee, J.S., Silverstein, R., Hoff, H.F., and Freeman, M.W. (2002). Scavenger receptors class A-I/II and CD36 are the principal receptors responsible for the uptake of modified low density lipoprotein leading to lipid loading in macrophages. The Journal of biological chemistry *277*, 49982-49988.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

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

Leblanc, S.K., Oates, C.W., and Raivio, T.L. (2011). Characterization of the induction and cellular role of the BaeSR two-component envelope stress response of Escherichia coli. Journal of bacteriology *193*, 3367-3375.

Lee, S., Byakwaga, H., Boum, Y., Burdo, T.H., Williams, K.C., Lederman, M.M., Huang, Y., Tracy, R.P., Cao, H., Haberer, J.E.*, et al.* (2017). Immunologic Pathways That Predict Mortality in HIV-Infected Ugandans Initiating Antiretroviral Therapy. J Infect Dis *215*, 1270-1274.

Lee, S.C., Chua, L.L., Yap, S.H., Khang, T.F., Leng, C.Y., Raja Azwa, R.I., Lewin, S.R., Kamarulzaman, A., Woo, Y.L., Lim, Y.A.L.*, et al.* (2018). Enrichment of gut-derived Fusobacterium is associated with suboptimal immune recovery in HIV-infected individuals. Scientific reports *8*, 14277.

Lewden, C., Chene, G., Morlat, P., Raffi, F., Dupon, M., Dellamonica, P., Pellegrin, J.-L., Katlama, C., Dabis, F., Leport, C*., et al.* (2007). HIV-infected adults with a CD4 cell count greater than 500 cells/mm3 on long-term combination antiretroviral therapy reach same mortality rates as the general population. Journal of acquired immune deficiency syndromes (1999) *46*, 72-77.

Ley, R., Turnbaugh, P., Klein, S., and Gordon, J. (2006a). Microbial ecology: human gut microbes associated with obesity. Nature *444*, 1022-1023.

Ley, R.E. (2016). Gut microbiota in 2015: Prevotella in the gut: choose carefully. Nature reviews Gastroenterology & hepatology *13*, 69-70.

Ley, R.E., Peterson, D.A., and Gordon, J.I. (2006b). Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell *124*, 837-848.

Li, Q., Duan, L., Estes, J.D., Ma, Z.-M., Rourke, T., Wang, Y., Reilly, C., Carlis, J., Miller, C.J., and Haase, A.T. (2005). Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells. Nature *434*, 1148-1152.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658-1659.

Lin, M.F., Lin, Y.Y., Yeh, H.W., and Lan, C.Y. (2014). Role of the BaeSR two-component system in the regulation of Acinetobacter baumannii adeAB genes and its correlation with tigecycline susceptibility. BMC microbiology *14*, 119.

Ling, Z., Jin, C., Xie, T., Cheng, Y., Li, L., and Wu, N. (2016). Alterations in the Fecal Microbiota of Patients with HIV-1 Infection: An Observational Study in A Chinese Population. Scientific reports *6*, 30673.

Littman, D.R., and Pamer, E.G. (2011). Role of the commensal microbiota in normal and pathogenic host immune responses. Cell host & microbe *10*, 311-323.

Liu, J., Johnson, R., Dillon, S., Kroehl, M., Frank, D.N., Tuncil, Y.E., Zhang, X., Ir, D., Robertson, C.E., Seifert, S.*, et al.* (2019). Among older adults, age-related changes in the stool microbiome differ by HIV-1 serostatus. EBioMedicine *40*, 583-594.

Livanos, A.E., Greiner, T.U., Vangay, P., Pathmasiri, W., Stewart, D., McRitchie, S., Li, H., Chung, J., Sohn, J., Kim, S.*, et al.* (2016). Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. Nature microbiology *1*, 16140.

Lohse, N., Hansen, A.-B.E., Pedersen, G., Kronborg, G., Gerstoft, J., Sørensen, H.T., Vaeth, M., and Obel, N. (2007). Survival of persons with and without HIV infection in Denmark, 1995-2005. Annals of internal medicine *146*, 87-95.

Louis, P., and Flint, H.J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS microbiology letters *294*, 1-8.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology *15*, 550.

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

Lozupone, C.A., Li, M., Campbell, T.B., Flores, S.C., Linderman, D., Gebert, M.J., Knight, R., Fontenot, A.P., and Palmer, B.E. (2013). Alterations in the gut microbiota associated with HIV-1 infection. Cell host & microbe *14*, 329-339.

Lozupone, C.a., Rhodes, M.E., Neff, C.P., Fontenot, A.P., Campbell, T.B., and Palmer, B.E. (2014). HIV-induced alteration in gut microbiota: driving factors, consequences, and effects of antiretroviral therapy. Gut microbes *5*, 562-570.

Lu, W., Feng, Y., Jing, F., Han, Y., Lyu, N., Liu, F., Li, J., Song, X., Xie, J., Qiu, Z.*, et al.* (2018). Association Between Gut Microbiota and CD4 Recovery in HIV-1 Infected Patients. Frontiers in microbiology *9*, 1451.

Lukens, J.R., Gurung, P., Vogel, P., Johnson, G.R., Carter, R.A., McGoldrick, D.J., Bandi, S.R., Calabrese, C.R., Vande Walle, L., Lamkanfi, M.*, et al.* (2014).

Dietary modulation of the microbiome affects autoinflammatory disease. Nature *516*, 246-249.

Luo, W., and Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics *29*, 1830-1831.

Lupp, C., Robertson, M.L., Wickham, M.E., Sekirov, I., Champion, O.L., Gaynor, E.C., and Finlay, B.B. (2007). Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. Cell Host Microbe *2*, 119-129.

Lyte, M. (2013). Microbial endocrinology in the microbiome-gut-brain axis: how bacterial production and utilization of neurochemicals influence behavior. PLoS Pathog *9*, e1003726.

Ma, C., Wu, X., Nawaz, M., Li, J., Yu, P., Moore, J.E., and Xu, J. (2011). Molecular characterization of fecal microbiota in patients with viral diarrhea. Current microbiology *63*, 259-266.

Maartens, G., Celum, C., and Lewin, S.R. (2014). HIV infection: epidemiology, pathogenesis, treatment, and prevention. Lancet *384*, 258-271.

Maloy, K.J., and Kullberg, M.C. (2008). IL-23 and Th17 cytokines in intestinal homeostasis. Mucosal Immunol *1*, 339-349.

Mankal, P.K., and Kotler, D.P. (2014). From wasting to obesity, changes in nutritional concerns in HIV/AIDS. Endocrinology and metabolism clinics of North America *43*, 647-663.

March, C., Regueiro, V., Llobet, E., Moranta, D., Morey, P., Garmendia, J., and Bengoechea, J.A. (2010). Dissection of host cell signal transduction during Acinetobacter baumannii-triggered inflammatory response. PLoS One *5*, e10033.

Marchetti, G., Tincati, C., and Silvestri, G. (2013). Microbial translocation in the pathogenesis of HIV infection and AIDS. Clinical microbiology reviews *26*, 2-18.

Marcus, J.L., Leyden, W.a., Chao, C.R., Chow, F.C., Horberg, M.a., Hurley, L.B., Klein, D.B., Quesenberry, C.P., Towner, W.J., and Silverberg, M.J. (2014). HIV

infection and incidence of ischemic stroke. AIDS (London, England) *28*, 1911-1919.

Marotz, C.A., Sanders, J.G., Zuniga, C., Zaramela, L.S., Knight, R., and Zengler, K. (2018). Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome *6*, 42.

Martinez, I., Stegen, J.C., Maldonado-Gomez, M.X., Eren, A.M., Siba, P.M., Greenhill, A.R., and Walter, J. (2015). The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. Cell reports *11*, 527-538.

Mattapallil, J.J., Douek, D.C., Hill, B., Nishimura, Y., Martin, M., and Roederer, M. (2005). Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. Nature *434*, 1093-1097.

Maurice, J.B., Garvey, L., Tsochatzis, E.A., Wiltshire, M., Cooke, G., Guppy, N., McDonald, J., Marchesi, J., Nelson, M., Kelleher, P.*, et al.* (2019). Monocyte-Macrophage activation is associated with NAFLD and liver fibrosis in HIV mono-infection independently of the gut microbiome and bacterial translocation. Aids.

Mavromanolakis, E., Maraki, S., Samonis, G., Tselentis, Y., and Cranidis, A. (1997). Effect of norfloxacin, trimethoprim-sulfamethoxazole and nitrofurantoin on fecal flora of women with recurrent urinary tract infections. Journal of chemotherapy *9*, 203-207.

Mazmanian, S.K., Round, J.L., and Kasper, D.L. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. Nature *453*, 620-625.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME journal *6*, 610-618.

McGuckin, M.A., Linden, S.K., Sutton, P., and Florin, T.H. (2011). Mucin dynamics and enteric pathogens. Nat Rev Microbiol *9*, 265-278.

McHardy, I.H., Li, X., Tong, M., Ruegger, P., Jacobs, J., Borneman, J., Anton, P., and Braun, J. (2013). HIV Infection is associated with compositional and functional shifts in the rectal mucosal microbiota. Microbiome *1*, 26.

McKibben, R.A., Margolick, J.B., Grinspoon, S., Li, X., Palella, F.J., Kingsley, L.A., Witt, M.D., George, R.T., Jacobson, L.P., Budoff, M.*, et al.* (2015). Elevated levels of monocyte activation markers are associated with subclinical atherosclerosis in men with and those without HIV infection. Journal of Infectious Diseases *211*, 1219-1228.

McMurdie, P.J., and Holmes, S. (2012). Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. Pac Symp Biocomput, 235-246.

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

Meehan, C.J., and Beiko, R.G. (2014). A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. Genome biology and evolution *6*, 703-713.

Mehandru, S., Poles, M.A., Tenner-Racz, K., Horowitz, A., Hurley, A., Hogan, C., Boden, D., Racz, P., and Markowitz, M. (2004). Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. The Journal of experimental medicine *200*, 761-770.

Meir, K.S., and Leitersdorf, E. (2004). Atherosclerosis in the apolipoprotein-E-deficient mouse: a decade of progress. Arterioscler Thromb Vasc Biol *24*, 1006-1014.

Mende, D.R., Sunagawa, S., Zeller, G., and Bork, P. (2013). Accurate and universal delineation of prokaryotic species. Nat Methods *10*, 881-884.

Metwaly, A., and Haller, D. (2019). Strain-Level Diversity in the Gut: The P. copri Case. Cell Host Microbe *25*, 349-350.

Milani, C., Ticinesi, A., Gerritsen, J., Nouvenne, A., Lugli, G.A., Mancabelli, L., Turroni, F., Duranti, S., Mangifesta, M., Viappiani, A.*, et al.* (2016). Gut microbiota composition and Clostridium difficile infection in hospitalized elderly individuals: a metagenomic study. Scientific reports *6*, 25945.

Miller, H., Ferris, R., and Phelps, B.R. (2016). The effect of probiotics on CD4 counts among people living with HIV: a systematic review. Beneficial microbes *7*, 345-351.

Mills, E.J., Bärnighausen, T., and Negin, J. (2012). HIV and aging--preparing for the challenges ahead. The New England journal of medicine *366*, 1270-1273.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: inter-individual variation and dynamic response to diet. Genome Research *21*, 1616-1625.

Monachese, M., Cunningham-Rundles, S., Diaz, M.A., Guerrant, R., Hummelen, R., Kemperman, R., Kerac, M., Kort, R., Merenstein, D., Panigrahi, P.*, et al.* (2011). Probiotics and prebiotics to combat enteric infections and HIV in the developing world: a consensus report. Gut microbes *2*, 198-207.

Monaco, C.L., Gootenberg, D.B., Zhao, G., Handley, S.A., Ghebremichael, M.S., Lim, E.S., Lankowski, A., Baldridge, M.T., Wilen, C.B., Flagg, M.*, et al.* (2016). Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell host & microbe *19*, 311-322.

Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., Reyes, J.A., Shah, S.A., LeLeiko, N., Snapper, S.B.*, et al.* (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome biology *13*, R79.

Muiru, A.N., Bibangambah, P., Hemphill, L., Sentongo, R., Kim, J.H., Triant, V.A., Bangsberg, D.R., Tsai, A.C., Martin, J.N., Haberer, J.E.*, et al.* (2018). Distribution and Performance of Cardiovascular Risk Scores in a Mixed Population of HIV-Infected and Community-Based HIV-Uninfected Individuals in Uganda. J Acquir Immune Defic Syndr *78*, 458-464.

Mutlu, E.A., Keshavarzian, A., Losurdo, J., Swanson, G., Siewe, B., Forsyth, C., French, A., Demarais, P., Sun, Y., Koenig, L.*, et al.* (2014). A compositional look at the human gastrointestinal microbiome and immune activation parameters in HIV infected subjects. PLoS pathogens *10*, e1003829.

Muyanja, D., Muzoora, C., Muyingo, A., Muyindike, W., and Siedner, M.J. (2015). High Prevalence of Metabolic Syndrome and Cardiovascular Disease Risk Among People with HIV on Stable ART in Southwestern Uganda. AIDS Patient Care STDS *30*, 4-10.

Nakano, M.M., Zuber, P., Glaser, P., Danchin, A., and Hulett, F.M. (1996). Two-component regulatory proteins ResD-ResE are required for transcriptional

activation of fnr upon oxygen limitation in Bacillus subtilis. Journal of bacteriology *178*, 3796-3802.

Nakashima, Y., Plump, A.S., Raines, E.W., Breslow, J.L., and Ross, R. (1994). ApoE-deficient mice develop lesions of all phases of atherosclerosis throughout the arterial tree. Arteriosclerosis and thrombosis : a journal of vascular biology *14*, 133-140.

Nayfach, S., Shi, Z.J., Seshadri, R., Pollard, K.S., and Kyrpides, N. (2019). Novel insights from uncultivated genomes of the global human gut microbiome. Nature.

Neff, C.P., Krueger, O., Xiong, K., Arif, S., Nusbacher, N., Schneider, J.M., Cunningham, A.W., Armstrong, A., Li, S., McCarter, M.D.*, et al.* (2018). Fecal Microbiota Composition Drives Immune Activation in HIV-infected Individuals. EBioMedicine *30*, 192-202.

Nelson, K.E., Weinstock, G.M., Highlander, S.K., Worley, K.C., Creasy, H.H., Wortman, J.R., Rusch, D.B., Mitreva, M., Sodergren, E., Chinwalla, A.T.*, et al.* (2010). A catalog of reference genomes from the human microbiome. Science *328*, 994-999.

Neves, A.L., Coelho, J., Couto, L., Leite-Moreira, A., and Roncon-Albuquerque, R. (2013). Metabolic endotoxemia: a molecular link between obesity and cardiovascular risk. Journal of molecular endocrinology *51*, R51-64.

Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and Pettersson, S. (2012). Host-gut microbiota metabolic interactions. Science *336*, 1262-1267.

Nobel, Y.R., Cox, L.M., Kirigin, F.F., Bokulich, N.A., Yamanishi, S., Teitler, I., Chung, J., Sohn, J., Barber, C.M., Goldfarb, D.S.*, et al.* (2015). Metabolic and metagenomic outcomes from early-life pulsed antibiotic treatment. Nature communications *6*, 7486.

Noguera-Julian, M., Rocafort, M., Guillen, Y., Rivera, J., Casadella, M., Nowak, P., Hildebrand, F., Zeller, G., Parera, M., Bellido, R.*, et al.* (2016). Gut Microbiota Linked to Sexual Preference and HIV Infection. EBioMedicine *5*, 135-146.

Norman, J.M., Handley, S.A., Baldridge, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., Fleshner, P.*, et al.* (2015). Disease-specific

alterations in the enteric virome in inflammatory bowel disease. Cell *160*, 447-460.

Norman, J.M., Handley, S.A., and Virgin, H.W. (2014). Kingdom-agnostic metagenomics and the importance of complete characterization of enteric microbial communities. Gastroenterology *146*, 1459-1469.

Nowak, P., Troseid, M., Avershina, E., Barqasho, B., Neogi, U., Holm, K., Hov, J.R., Noyan, K., Vesterbacka, J., Svärd, J.*, et al.* (2015). Gut microbiota diversity predicts immune status in HIV-1 infection. AIDS (London, England) *29*, 2409-2418.

Nowak, R.G., Bentzen, S.M., Ravel, J., Crowell, T.A., Dauda, W., Ma, B., Liu, H., Blattner, W.A., Baral, S.D., Charurat, M.E.*, et al.* (2019). Anal Microbial Patterns and Oncogenic Human Papillomavirus in a Pilot Study of Nigerian Men Who Have Sex with Men at Risk for or Living with HIV. AIDS research and human retroviruses *35*, 267-275.

Nowak, R.G., Bentzen, S.M., Ravel, J., Crowell, T.A., Dauda, W., Ma, B., Liu, H., Blattner, W.A., Baral, S.D., Charurat, M.E.*, et al.* (2017). Rectal microbiota among HIV-uninfected, untreated HIV, and treated HIV-infected in Nigeria. Aids *31*, 857-862.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., Prjibelski, A.D., Pyshkin, A., Sirotkin, A., Sirotkin, Y.*, et al.* (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. Journal of computational biology : a journal of computational molecular cell biology *20*, 714-737.

Obregon-Tito, A.J., Tito, R.Y., Metcalf, J., Sankaranarayanan, K., Clemente, J.C., Ursell, L.K., Zech Xu, Z., Van Treuren, W., Knight, R., Gaffney, P.M.*, et al.* (2015). Subsistence strategies in traditional societies distinguish gut microbiomes. Nature communications *6*, 6505.

Okatch, H., Bellamy, S.L., Han, X., Ratshaa, B., Steenhoff, A.P., Mosepele, M., Bisson, G.P., and Gross, R. (2016). CD4 Cell Counts at Antiretroviral Therapy Initiation in Botswana Have Been Increasing. Clin Infect Dis *62*, 669-670.

Okello, S., Kanyesigye, M., Muyindike, W.R., Annex, B.H., Hunt, P.W., Haneuse, S., and Siedner, M.J. (2015). Incidence and predictors of hypertension in adults with

HIV-initiating antiretroviral therapy in south-western Uganda. Journal of hypertension *33*, 2039-2045.

Okoye, A.A., and Picker, L.J. (2013). CD4(+) T-cell depletion in HIV infection: mechanisms of immunological failure. Immunological reviews *254*, 54-64.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H.*, et al.* (2013). vegan: Community Ecology Package. R package version 2.0-10.

Oldenburg, C.E., Sie, A., Coulibaly, B., Ouermi, L., Dah, C., Tapsoba, C., Barnighausen, T., Ray, K.J., Zhong, L., Cummings, S.*, et al.* (2018). Effect of Commonly Used Pediatric Antibiotics on Gut Microbial Diversity in Preschool Children in Burkina Faso: A Randomized Clinical Trial. Open forum infectious diseases *5*, ofy289.

Palermo, B., Bosch, R.J., Bennett, K., and Jacobson, J.M. (2011). Body mass index and CD4+ T-lymphocyte recovery in HIV-infected men with viral suppression on antiretroviral therapy. HIV clinical trials *12*, 222-227.

Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526-528.

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P.*, et al.* (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell *176*, 649-662.e620.

Peleg, A.Y., de Breij, A., Adams, M.D., Cerqueira, G.M., Mocali, S., Galardini, M., Nibbering, P.H., Earl, A.M., Ward, D.V., Paterson, D.L.*, et al.* (2012). The success of acinetobacter species; genetic, metabolic and virulence attributes. PLoS One *7*, e46984.

Pescatore, N.A., Pollak, R., Kraft, C.S., Mulle, J.G., and Kelley, C.F. (2018). Short Communication: Anatomic Site of Sampling and the Rectal Mucosal Microbiota in HIV Negative Men Who Have Sex with Men Engaging in Condomless Receptive Anal Intercourse. AIDS research and human retroviruses *34*, 277-281.

Peterson, D.A., and Cardona, R.A. (2010). Specificity of the adaptive immune response to the gut microbiota. Adv Immunol *107*, 71-107.

Peterson, D.A., McNulty, N.P., Guruge, J.L., and Gordon, J.I. (2007). IgA response to symbiotic bacteria as a mediator of gut homeostasis. Cell Host Microbe *2*, 328-339.

Peterson, D.A., and Turnbaugh, P.J. (2010). A microbe-dependent viral key to Crohn's box. Sci Transl Med *2*, 43ps39.

Phillips, A.N., Lampe, F.C., Smith, C.J., Geretti, A.M., Rodger, A., Lodwick, R.K., Cambiano, V., Tsintas, R., and Johnson, M.A. (2010). Ongoing changes in HIV RNA levels during untreated HIV infection: implications for CD4 cell count depletion. Aids *24*, 1561-1567.

Pickard, J.M., Maurice, C.F., Kinnebrew, M.A., Abt, M.C., Schenten, D., Golovkina, T.V., Bogatyrev, S.R., Ismagilov, R.F., Pamer, E.G., Turnbaugh, P.J.*, et al.* (2014). Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. Nature *514*, 638-641.

Pinto-Cardoso, S., Lozupone, C., Briceno, O., Alva-Hernandez, S., Tellez, N., Adriana, A., Murakami-Ogasawara, A., and Reyes-Teran, G. (2017). Fecal Bacterial Communities in treated HIV infected individuals on two antiretroviral regimens. Scientific reports *7*, 43741.

Png, C.W., Lindén, S.K., Gilshenan, K.S., Zoetendal, E.G., McSweeney, C.S., Sly, L.I., McGuckin, M.a., and Florin, T.H.J. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. The American journal of gastroenterology *105*, 2420-2428.

Popova, M., Martin, C., and Morgavi, D.P. (2010). Improved protocol for high-quality co-extraction of DNA and RNA from rumen digesta. Folia microbiologica *55*, 368-372.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490.

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

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D.*, et al.* (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55-60.

Qing, Y., Xie, H., Su, C., Wang, Y., Yu, Q., Pang, Q., and Cui, F. (2018). Gut Microbiome, Short-Chain Fatty Acids, and Mucosa Injury in Young Adults with Human Immunodeficiency Virus Infection. Digestive diseases and sciences.

Rajilic-Stojanovic, M., Shanahan, F., Guarner, F., and de Vos, W.M. (2013). Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. Inflammatory bowel diseases *19*, 481-488.

Ranoa, D.R., Kelley, S.L., and Tapping, R.I. (2013). Human lipopolysaccharide-binding protein (LBP) and CD14 independently deliver triacylated lipoproteins to Toll-like receptor 1 (TLR1) and TLR2 and enhance formation of the ternary signaling complex. The Journal of biological chemistry *288*, 9729-9741.

Reid, G. (2010). The potential role for probiotic yogurt for people living with HIV/AIDS. Gut microbes *1*, 411-414.

Reid, M.J.A., Mosepele, M., Tsima, B.M., and Gross, R. (2012). Addressing the challenge of the emerging NCD epidemic: lessons learned from Botswana's response to the HIV epidemic. Public health action *2*, 47-49.

Reimund, J.M., Wittersheim, C., Dumont, S., Muller, C.D., Baumann, R., Poindron, P., and Duclos, B. (1996). Mucosal inflammatory cytokine production by intestinal biopsies in patients with ulcerative colitis and Crohn's disease. Journal of clinical immunology *16*, 144-150.

Renshaw, M.A., Olds, B.P., Jerde, C.L., McVeigh, M.M., and Lodge, D.M. (2015). The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol-chloroform-isoamyl alcohol DNA extraction. Molecular ecology resources *15*, 168-176.

Reus, S., Portilla, J., Sanchez-Paya, J., Giner, L., Frances, R., Such, J., Boix, V., Merino, E., and Gimeno, A. (2013). Low-level HIV viremia is associated with microbial translocation and inflammation. J Acquir Immune Defic Syndr *62*, 129-134.

Rey, F.E., Faith, J.J., Bain, J., Muehlbauer, M.J., Stevens, R.D., Newgard, C.B., and Gordon, J.I. (2010). Dissecting the in vivo metabolic potential of two human gut acetogens. The Journal of biological chemistry *285*, 22082-22090.

Rey, F.E., Gonzalez, M.D., Cheng, J., Wu, M., Ahern, P.P., and Gordon, J.I. (2013). Metabolic niche of a prominent sulfate-reducing human gut bacterium. Proc Natl Acad Sci U S A *110*, 13582-13587.

Reyes, A., Wu, M., McNulty, N.P., Rohwer, F.L., and Gordon, J.I. (2013). Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. Proceedings of the National Academy of Sciences of the United States of America *110*, 20236-20241.

Rogler, G., and Andus, T. (1998). Cytokines in inflammatory bowel disease. World journal of surgery *22*, 382-389.

Roser, M., and Ritchie, H. (2019). HIV / AIDS. In Our World in Data.

Ross, A.C., Rizk, N., O'Riordan, M.A., Dogra, V., El-Bejjani, D., Storer, N., Harrill, D., Tungsiripat, M., Adell, J., and McComsey, G.a. (2009). Relationship between inflammatory markers, endothelial activation markers, and carotid intima-media thickness in HIV-infected patients receiving antiretroviral therapy. Clin Infect Dis *49*, 1119-1127.

Round, J.L., and Mazmanian, S.K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. Proc Natl Acad Sci U S A *107*, 12204-12209.

Ruiz, V.E., Battaglia, T., Kurtz, Z.D., Bijnens, L., Ou, A., Engstrand, I., Zheng, X., Iizumi, T., Mullins, B.J., Muller, C.L.*, et al.* (2017). A single early-in-life macrolide course has lasting effects on murine microbial network topology and immunity. Nature communications *8*, 518.

Salah Ud-Din, A.I.M., and Roujeinikova, A. (2017). Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. Cellular and molecular life sciences : CMLS *74*, 3293-3303.

San-Juan-Vergara, H., Zurek, E., Ajami, N.J., Mogollon, C., Pena, M., Portnoy, I., Velez, J.I., Cadena-Cruz, C., Diaz-Olmos, Y., Hurtado-Gomez, L.*, et al.* (2018). A

Lachnospiraceae-dominated bacterial signature in the fecal microbiota of HIV-infected individuals from Colombia, South America. Scientific reports *8*, 4479.

Sandler, N.G., and Douek, D.C. (2012). Microbial translocation in HIV infection: causes, consequences and treatment opportunities. Nature reviews Microbiology *10*, 655-666.

Sandler, N.G., Wand, H., Roque, A., Law, M., Nason, M.C., Nixon, D.E., Pedersen, C., Ruxrungtham, K., Lewin, S.R., Emery, S.*, et al.* (2011). Plasma levels of soluble CD14 independently predict mortality in HIV infection. J Infect Dis *203*, 780-790.

Sangster, W., Hegarty, J.P., Schieffer, K.M., Wright, J.R., Hackman, J., Toole, D.R., Lamendella, R., and Stewart, D.B., Sr. (2016). Bacterial and Fungal Microbiota Changes Distinguish C. difficile Infection from Other Forms of Diarrhea: Results of a Prospective Inpatient Study. Frontiers in microbiology *7*, 789.

Sarbini, S.R., Kolida, S., Naeye, T., Einerhand, A., Brison, Y., Remaud-Simeon, M., Monsan, P., Gibson, G.R., and Rastall, R.A. (2011). In vitro fermentation of linear and alpha-1,2-branched dextrans by the human fecal microbiota. Applied and environmental microbiology *77*, 5307-5315.

Satokari, R., Fuentes, S., Mattila, E., Jalanka, J., de Vos, W.M., and Arkkila, P. (2014). Fecal transplantation treatment of antibiotic-induced, noninfectious colitis and long-term microbiota follow-up. Case Rep Med *2014*, 913867.

Scher, J.U., Sczesnak, A., Longman, R.S., Segata, N., Ubeda, C., Bielski, C., Rostron, T., Cerundolo, V., Pamer, E.G., Abramson, S.B.*, et al.* (2013). Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. eLife *2*, e01202.

Scher, J.U., Ubeda, C., Artacho, A., Attur, M., Isaac, S., Reddy, S.M., Marmon, S., Neimann, A., Brusca, S., Patel, T.*, et al.* (2015). Decreased bacterial diversity characterizes the altered gut microbiota in patients with psoriatic arthritis, resembling dysbiosis in inflammatory bowel disease. Arthritis & rheumatology *67*, 128-139.

Schledzewski, K., Falkowski, M., Moldenhauer, G., Metharom, P., Kzhyshkowska, J., Ganss, R., Demory, A., Falkowska-Hansen, B., Kurzen, H., Ugurel, S.*, et al.* (2006). Lymphatic endothelium-specific hyaluronan receptor LYVE-1 is expressed by stabilin-1+, F4/80+, CD11b+ macrophages in malignant tumours and wound healing tissue in vivo and in bone marrow cultures in vitro:

implications for the assessment of lymphangiogenesis. The Journal of pathology *209*, 67-77.

Schneider, T., Ullrich, R., and Zeitz, M. (1996). The immunologic aspects of human immunodeficiency virus infection in the gastrointestinal tract. Seminars in gastrointestinal disease *7*, 19-29.

Schnorr, S.L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M.*, et al.* (2014). Gut microbiome of the Hadza hunter-gatherers. Nature communications *5*, 3654.

Schubert, A.M., Rogers, M.A., Ring, C., Mogle, J., Petrosino, J.P., Young, V.B., Aronoff, D.M., and Schloss, P.D. (2014). Microbiome data distinguish patients with Clostridium difficile infection and non-C. difficile-associated diarrhea from healthy controls. mBio *5*, e01021-01014.

Schubert, A.M., Sinani, H., and Schloss, P.D. (2015). Antibiotic-Induced Alterations of the Murine Gut Microbiota and Subsequent Effects on Colonization Resistance against Clostridium difficile. mBio *6*, e00974.

Schuetz, A., Deleage, C., Sereti, I., Rerknimitr, R., Phanuphak, N., Phuang-Ngern, Y., Estes, J.D., Sandler, N.G., Sukhumvittaya, S., Marovich, M.*, et al.* (2014). Initiation of ART during early acute HIV infection preserves mucosal Th17 function and reverses HIV-related immune activation. PLoS Pathog *10*, e1004543.

Schunter, M., Chu, H., Hayes, T.L., McConnell, D., Crawford, S.S., Luciw, P.A., Bengmark, S., Asmuth, D.M., Brown, J., Bevins, C.L.*, et al.* (2012). Randomized pilot trial of a synbiotic dietary supplement in chronic HIV-1 infection. BMC complementary and alternative medicine *12*, 84.

Scott, K.P., Antoine, J.M., Midtvedt, T., and van Hemert, S. (2015). Manipulating the gut microbiota to maintain health and treat disease. Microb Ecol Health Dis *26*, 25877.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods *9*, 811-814.

Selik, R.M., Mokotoff, E.D., Branson, B., Owen, S.M., Whitmore, S., and Hall, H.I. (2014). Revised surveillance case definition for HIV infection—United States, 2014. MMWR *63*, 1-10.

Serrano-Villar, S., Vazquez-Castellanos, J.F., Vallejo, A., Latorre, A., Sainz, T., Ferrando-Martinez, S., Rojo, D., Martinez-Botas, J., Del Romero, J., Madrid, N.*, et al.* (2017). The effects of prebiotics on microbial dysbiosis, butyrate production and immunity in HIV-infected subjects. Mucosal Immunol *10*, 1279-1293.

Sessa, L., Reddel, S., Manno, E., Quagliariello, A., Cotugno, N., Del Chierico, F., Amodio, D., Capponi, C., Leone, F., Bernardi, S.*, et al.* (2019). Distinct gut microbiota profile in ART-treated perinatally HIV-infected patients associated with cardiac and inflammatory biomarkers. Aids.

Shaw, K.A., Bertha, M., Hofmekler, T., Chopra, P., Vatanen, T., Srivatsa, A., Prince, J., Kumar, A., Sauer, C., Zwick, M.E.*, et al.* (2016). Dysbiosis, inflammation, and response to treatment: a longitudinal study of pediatric subjects with newly diagnosed inflammatory bowel disease. Genome medicine *8*, 75.

Siedner, M.J., Kim, J.-H., Nakku, R.S., Bibangambah, P., Hemphill, L., Triant, V.A., Haberer, J.E., Martin, J.N., Mocello, A.R., Boum, Y.*, et al.* (2016a). Persistent Immune Activation and Carotid Atherosclerosis in HIV-Infected Ugandans Receiving Antiretroviral Therapy. The Journal of infectious diseases *213*, 370-378.

Siedner, M.J., Kim, J.H., Nakku, R.S., Hemphill, L., Triant, V.A., Haberer, J.E., Martin, J.N., Boum, Y., 2nd, Kwon, D.S., Tsai, A.C.*, et al.* (2016b). HIV infection and arterial stiffness among older-adults taking antiretroviral therapy in rural Uganda. Aids *30*, 667-670.

Siedner, M.J., Lankowski, A., Tsai, A.C., Muzoora, C., Martin, J.N., Hunt, P.W., Haberer, J.E., and Bangsberg, D.R. (2013). GPS-measured distance to clinic, but not self-reported transportation factors, are associated with missed HIV clinic visits in rural Uganda. Aids *27*, 1503-1508.

Sievers, F., and Higgins, D.G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. Methods in molecular biology *1079*, 105-116.

Sinha, R., Ahn, J., Sampson, J.N., Shi, J., Yu, G., Xiong, X., Hayes, R.B., and Goedert, J.J. (2016). Fecal Microbiota, Fecal Metabolome, and Colorectal Cancer Interrelations. PLoS One *11*, e0152126.

Smits, S.A., Leach, J., Sonnenburg, E.D., Gonzalez, C.G., Lichtman, J.S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J.E.*, et al.* (2017). Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science *357*, 802-806.

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

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312-1313.

Stamey, T.A., Condy, M., and Mihara, G. (1977). Prophylactic efficacy of nitrofurantoin macrocrystals and trimethoprim-sulfamethoxazole in urinary infections. Biologic effects on the vaginal and rectal flora. N Engl J Med *296*, 780-783.

Stearns, J.C., Lynch, M.D., Senadheera, D.B., Tenenbaum, H.C., Goldberg, M.B., Cvitkovitch, D.G., Croitoru, K., Moreno-Hagelsieb, G., and Neufeld, J.D. (2011). Bacterial biogeography of the human digestive tract. Scientific reports *1*, 170.

Stiksrud, B., Nowak, P., Nwosu, F.C., Kvale, D., Thalme, A., Sonnerborg, A., Ueland, P.M., Holm, K., Birkeland, S.E., Dahm, A.E.*, et al.* (2015). Reduced Levels of D-dimer and Changes in Gut Microbiota Composition After Probiotic Intervention in HIV-Infected Individuals on Stable ART. J Acquir Immune Defic Syndr *70*, 329-337.

Subramanian, S., Tawakol, A., Burdo, T.H., Abbara, S., Wei, J., Vijayakumar, J., Corsini, E., Abdelbaky, A., Zanni, M.V., Hoffmann, U.*, et al.* (2012). Arterial inflammation in patients with HIV. JAMA : the journal of the American Medical Association *308*, 379-386.

Sun, Y., Ma, Y., Lin, P., Tang, Y.W., Yang, L., Shen, Y., Zhang, R., Liu, L., Cheng, J., Shao, J.*, et al.* (2016). Fecal bacterial microbiome diversity in chronic HIV-infected patients in China. Emerging microbes & infections *5*, e31.

Sze, M.A., and Schloss, P.D. (2016). Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. mBio *7*.

Tannock, G.W., Lawley, B., Munro, K., Lay, C., Taylor, C., Daynes, C., Baladjay, L., McLeod, R., and Thompson-Fawcett, M. (2012). Comprehensive analysis of the

bacterial content of stool from patients with chronic pouchitis, normal pouches, or familial adenomatous polyposis pouches. Inflammatory bowel diseases *18*, 925-934.

Tanoue, T., Atarashi, K., and Honda, K. (2016). Development and maintenance of intestinal regulatory T cells. Nature reviews Immunology *16*, 295-309.

Tanoue, T., Nishitani, Y., Kanazawa, K., Hashimoto, T., and Mizuno, M. (2008). In vitro model to estimate gut inflammation using co-cultured Caco-2 and RAW264.7 cells. Biochemical and biophysical research communications *374*, 565-569.

Tapping, R.I., and Tobias, P.S. (2000). Soluble CD14-mediated cellular responses to lipopolysaccharide. Chemical immunology *74*, 108-121.

Team, R.C. (2013). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing).

Tenenbaum, D. (2018). KEGGREST: Client-side REST access to KEGG.

Thaiss, C.A., Zmora, N., Levy, M., and Elinav, E. (2016). The microbiome and innate immunity. Nature *535*, 65-74.

Thomas, C.J., Kapoor, M., Sharma, S., Bausinger, H., Zyilan, U., Lipsker, D., Hanau, D., and Surolia, A. (2002). Evidence of a trimolecular complex involving LPS, LPS binding protein and soluble CD14 as an effector of LPS response. FEBS letters *531*, 184-188.

Tien, P.C., Choi, A.I., Zolopa, A.R., Benson, C., Tracy, R., Scherzer, R., Bacchetti, P., Shlipak, M., and Grunfeld, C. (2010). Inflammation and mortality in HIV-infected adults: analysis of the FRAM study cohort. Journal of acquired immune deficiency syndromes (1999) *55*, 316-322.

Torres, B., Guardo, A.C., Leal, L., Leon, A., Lucero, C., Alvarez-Martinez, M.J., Martinez, M.J., Vila, J., Martinez-Rebollar, M., Gonzalez-Cordon, A.*, et al.* (2014). Protease inhibitor monotherapy is associated with a higher level of monocyte activation, bacterial translocation and inflammation. Journal of the International AIDS Society *17*, 19246.

Triant, V.A., Brown, T.T., Lee, H., and Grinspoon, S.K. (2008). Fracture prevalence among human immunodeficiency virus (HIV)-infected versus non-HIV-infected patients in a large U.S. healthcare system. The Journal of clinical endocrinology and metabolism *93*, 3499-3504.

Triant, V.a., and Grinspoon, S.K. (2011). Immune dysregulation and vascular risk in HIV-infected patients: implications for clinical care. The Journal of infectious diseases *203*, 439-441.

Triant, V.A., Lee, H., Hadigan, C., and Grinspoon, S.K. (2007). Increased acute myocardial infarction rates and cardiovascular risk factors among patients with human immunodeficiency virus disease. The Journal of clinical endocrinology and metabolism *92*, 2506-2512.

Triant, V.A., Meigs, J.B., and Grinspoon, S.K. (2009). Association of C-reactive protein and HIV infection with acute myocardial infarction. JAcquirImmuneDeficSyndr *51*, 268-273.

Troy, E.B., and Kasper, D.L. (2010). Beneficial effects of Bacteroides fragilis polysaccharides on the immune system. Frontiers in bioscience *15*, 25-34.

Truong, D.T., Tett, A., Pasolli, E., Huttenhower, C., and Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res *27*, 626-638.

Turnbaugh, P.J., Backhed, F., Fulton, L., and Gordon, J.I. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe *3*, 213-223.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007). The human microbiome project. Nature *449*, 804-810.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.a., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. Nature *444*, 1027-1031.

Turroni, S., Rampelli, S., Centanni, M., Schnorr, S.L., Consolandi, C., Severgnini, M., Peano, C., Soverini, M., Falconi, M., Crittenden, A.N.*, et al.* (2016). Enterocyte-Associated Microbiome of the Hadza Hunter-Gatherers. Frontiers in microbiology *7*, 865.

Ubeda, C., Bucci, V., Caballero, S., Djukovic, A., Toussaint, N.C., Equinda, M., Lipuma, L., Ling, L., Gobourne, A., No, D.*, et al.* (2013). Intestinal microbiota containing Barnesiella species cures vancomycin-resistant Enterococcus faecium colonization. Infection and immunity *81*, 965-973.

UNAIDS (2018). UNAIDS data 2018. Unaids, 376.

Vaishnava, S., Yamamoto, M., Severson, K.M., Ruhn, K.A., Yu, X., Koren, O., Ley, R., Wakeland, E.K., and Hooper, L.V. (2011). The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. Science *334*, 255-258.

Van De Walle, J., Hendrickx, A., Romier, B., Larondelle, Y., and Schneider, Y.J. (2010). Inflammatory parameters in Caco-2 cells: effect of stimuli nature, concentration, combination and cell differentiation. Toxicology in vitro : an international journal published in association with BIBRA *24*, 1441-1449.

Vangay, P., Johnson, A.J., Ward, T.L., Al-Ghalith, G.A., Shields-Cutler, R.R., Hillmann, B.M., Lucas, S.K., Beura, L.K., Thompson, E.A., Till, L.M.*, et al.* (2018). US Immigration Westernizes the Human Gut Microbiome. Cell *175*, 962-972 e910.

Vannucci, L., Stepankova, R., Kozakova, H., Fiserova, A., Rossmann, P., and Tlaskalova-Hogenova, H. (2008). Colorectal carcinogenesis in germ-free and conventionally reared rats: different intestinal environments affect the systemic immunity. Int J Oncol *32*, 609-617.

Vatanen, T., Kostic, A.D., d'Hennezel, E., Siljander, H., Franzosa, E.A., Yassour, M., Kolde, R., Vlamakis, H., Arthur, T.D., Hamalainen, A.M.*, et al.* (2016). Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. Cell *165*, 842-853.

Vazquez-Castellanos, J.F., Serrano-Villar, S., Jimenez-Hernandez, N., Soto Del Rio, M.D., Gayo, S., Rojo, D., Ferrer, M., Barbas, C., Moreno, S., Estrada, V.*, et al.* (2018). Interplay between gut microbiota metabolism and inflammation in HIV infection. The ISME journal *12*, 1964-1976.

Vazquez-Castellanos, J.F., Serrano-Villar, S., Latorre, A., Artacho, A., Ferrus, M.L., Madrid, N., Vallejo, A., Sainz, T., Martinez-Botas, J., Ferrando-Martinez, S.*, et al.* (2014). Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals. Mucosal Immunol.

Vázquez-Castellanos, J.F., Serrano-Villar, S., Latorre, A., Artacho, A., Ferrús, M.L., Madrid, N., Vallejo, A., Sainz, T., Martínez-Botas, J., Ferrando-Martínez, S*., et al.* (2015). Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals. Mucosal immunology *8*, 760-772.

Veazey, R.S., DeMaria, M., Chalifoux, L.V., Shvetz, D.E., Pauley, D.R., Knight, H.L., Rosenzweig, M., Johnson, R.P., Desrosiers, R.C., and Lackner, A.A. (1998). Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. Science (New York, NY) *280*, 427-431.

Vesterbacka, J., Rivera, J., Noyan, K., Parera, M., Neogi, U., Calle, M., Paredes, R., Sonnerborg, A., Noguera-Julian, M., and Nowak, P. (2017). Richer gut microbiota with distinct metabolic profile in HIV infected Elite Controllers. Scientific reports *7*, 6269.

Vijay-Kumar, M., Aitken, J.D., Carvalho, F.A., Cullender, T.C., Mwangi, S., Srinivasan, S., Sitaraman, S.V., Knight, R., Ley, R.E., and Gewirtz, A.T. (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. In Science, pp. 228-231.

Villanueva-Millan, M.J., Perez-Matute, P., Recio-Fernandez, E., Lezana Rosales, J.M., and Oteo, J.A. (2017). Differential effects of antiretrovirals on microbial translocation and gut microbiota composition of HIV-infected patients. Journal of the International AIDS Society *20*, 21526.

Villar-Garcia, J., Hernandez, J.J., Guerri-Fernandez, R., Gonzalez, A., Lerma, E., Guelar, A., Saenz, D., Sorli, L., Montero, M., Horcajada, J.P*., et al.* (2015). Effect of probiotics (Saccharomyces boulardii) on microbial translocation and inflammation in HIV-treated patients: a double-blind, randomized, placebo-controlled trial. J Acquir Immune Defic Syndr *68*, 256-263.

Virgin, H.W. (2014). The virome in mammalian physiology and disease. Cell *157*, 142-150.

Virgin, H.W., and Todd, J.A. (2011). Metagenomics and personalized medicine. Cell *147*, 44-56.

Virgin, H.W., Wherry, E.J., and Ahmed, R. (2009). Redefining chronic viral infection. Cell *138*, 30-50.

Vogtmann, E., Hua, X., Zeller, G., Sunagawa, S., Voigt, A.Y., Hercog, R., Goedert, J.J., Shi, J., Bork, P., and Sinha, R. (2016). Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. PLoS One *11*, e0155362.

Vujkovic-Cvijin, I., Dunham, R.M., Iwai, S., Maher, M.C., Albright, R.G., Broadhurst, M.J., Hernandez, R.D., Lederman, M.M., Huang, Y., Somsouk, M.*, et al.* (2013). Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism. Science translational medicine *5*, 193ra191.

Walker, R.I., Caldwell, M.B., Lee, E.C., Guerry, P., Trust, T.J., and Ruiz-Palacios, G.M. (1986). Pathophysiology of Campylobacter enteritis. Microbiological reviews *50*, 81-94.

Wang, J., Qi, J., Zhao, H., He, S., Zhang, Y., Wei, S., and Zhao, F. (2013). Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. Scientific reports *3*, 1843.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology *73*, 5261-5267.

Wang, T., Cai, G., Qiu, Y., Fei, N., Zhang, M., Pang, X., Jia, W., Cai, S., and Zhao, L. (2012). Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. The ISME journal *6*, 320-329.

Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., Levison, B.S., Dugar, B., Feldstein, A.E., Britt, E.B., Fu, X., Chung, Y.-M.*, et al.* (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature *472*, 57-63.

Wang, Z., Roberts, A.B., Buffa, J.A., Levison, B.S., Zhu, W., Org, E., Gu, X., Huang, Y., Zamanian-Daryoush, M., Culley, M.K.*, et al.* (2015). Non-lethal Inhibition of Gut Microbial Trimethylamine Production for the Treatment of Atherosclerosis. Cell *163*, 1585-1595.

Weingarden, A., Gonzalez, A., Vazquez-Baeza, Y., Weiss, S., Humphry, G., Berg-Lyons, D., Knights, D., Unno, T., Bobr, A., Kang, J.*, et al.* (2015). Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent Clostridium difficile infection. Microbiome *3*, 10.

Wen, L., Ley, R., Volchkov, P., Stranges, P., Avanesyan, L., Stonebraker, A., Hu, C., Wong, F., Szot, G., and Bluestone, J. (2008). Innate immunity and intestinal microbiota in the development of Type 1 diabetes. In Nature, pp. 1109-1113.

Williams, B., Weber, K., Chlipala, G., Evans, C., Morack, R., and French, A. (2019). HIV Status Does Not Affect Rectal Microbiome Composition, Diversity, or Stability over Time: A Chicago Women's Interagency HIV Study. AIDS research and human retroviruses *35*, 260-266.

Willing, B.P., Dicksved, J., Halfvarson, J., Andersson, A.F., Lucio, M., Zheng, Z., Jarnerot, G., Tysk, C., Jansson, J.K., and Engstrand, L. (2010). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. Gastroenterology *139*, 1844-1854 e1841.

Winglee, K., Howard, A.G., Sha, W., Gharaibeh, R.Z., Liu, J., Jin, D., Fodor, A.A., and Gordon-Larsen, P. (2017). Recent urbanization in China is correlated with a Westernized microbiome encoding increased virulence and antibiotic resistance genes. Microbiome *5*, 121.

Wlodarska, M., Willing, B., Keeney, K.M., Menendez, A., Bergstrom, K.S., Gill, N., Russell, S.L., Vallance, B.A., and Finlay, B.B. (2011). Antibiotic treatment alters the colonic mucus layer and predisposes the host to exacerbated Citrobacter rodentium-induced colitis. Infection and immunity *79*, 1536-1545.

World Health Organization (2006). Guidelines on co-trimoxazole prophylaxis for HIV-related infections among children, adolescents and adults: Recommendations for a public health approach (Geneva, Switzerland: World Health Organization Press).

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S.a., Bewtra, M., Knights, D., Walters, W.a., Knight, R.*, et al.* (2011). Linking long-term dietary patterns with gut microbial enterotypes. Science (New York, NY) *334*, 105-108.

Wylie, K.M., Weinstock, G.M., and Storch, G.A. (2012). Emerging view of the human virome. Translational research : the journal of laboratory and clinical medicine *160*, 283-290.

Yang, F., Yang, J., Zhang, X., Chen, L., Jiang, Y., Yan, Y., Tang, X., Wang, J., Xiong, Z., Dong, J.*, et al.* (2005). Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. Nucleic Acids Res *33*, 6445-6458.

Yang, L., Poles, M.A., Fisch, G.S., Ma, Y., Nossa, C., Phelan, J.A., and Pei, Z. (2016). HIV-induced immunosuppression is associated with colonization of the proximal gut by environmental bacteria. Aids *30*, 19-29.

Yasuda, K., Oh, K., Ren, B., Tickle, T.L., Franzosa, E.A., Wachtman, L.M., Miller, A.D., Westmoreland, S.V., Mansfield, K.G., Vallender, E.J.*, et al.* (2015). Biogeography of the intestinal mucosal and lumenal microbiome in the rhesus macaque. Cell Host Microbe *17*, 385-391.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P.*, et al.* (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222-227.

Yu, B., and Wright, S.D. (1995). LPS-dependent interaction of Mac-2-binding protein with immobilized CD14. Journal of inflammation *45*, 115-125.

Yu, G., Fadrosh, D., Ma, B., Ravel, J., and Goedert, J.J. (2014). Anal microbiota profiles in HIV-positive and HIV-negative MSM. AIDS (London, England) *28*, 753-760.

Yuan, H., Zelkha, S., Zelka, S., Burkatovskaya, M., Gupte, R., Leeman, S.E., and Amar, S. (2013). Pivotal role of NOD2 in inflammatory processes affecting atherosclerosis and periodontal bone loss. Proceedings of the National Academy of Sciences of the United States of America *110*, E5059-5068.

Zackular, J.P., Baxter, N.T., Iverson, K.D., Sadler, W.D., Petrosino, J.F., Chen, G.Y., and Schloss, P.D. (2013). The gut microbiome modulates colon tumorigenesis. mBio *4*, e00692-00613.

Ze, X., Duncan, S.H., Louis, P., and Flint, H.J. (2012). Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. The ISME journal *6*, 1535-1543.

Zevin, A.S., McKinnon, L., Burgener, A., and Klatt, N.R. (2016). Microbial translocation and microbiome dysbiosis in HIV-associated immune activation. Current opinion in HIV and AIDS *11*, 182-190.

Zhang, H., and Luo, X.M. (2015). Control of commensal microbiota by the adaptive immune system. Gut microbes *6*, 156-160.

Zhou, X., Zhang, N., Xia, L., Li, Q., Shao, J., Shen, Q., and Zhang, R. (2018a). ResDE Two-Component Regulatory System Mediates Oxygen Limitation-Induced Biofilm Formation by Bacillus amyloliquefaciens SQR9. Applied and environmental microbiology *84*.

Zhou, Y., Ou, Z., Tang, X., Zhou, Y., Xu, H., Wang, X., Li, K., He, J., Du, Y., Wang, H*., et al.* (2018b). Alterations in the gut microbiota of patients with acquired immune deficiency syndrome. Journal of cellular and molecular medicine *22*, 2263-2271.

Zhu, W., Gregory, J.C., Org, E., Buffa, J.A., Gupta, N., Wang, Z., Li, L., Fu, X., Wu, Y., Mehrabian, M*., et al.* (2016). Gut Microbial Metabolite TMAO Enhances Platelet Hyperreactivity and Thrombosis Risk. Cell *165*, 111-124.

Zietak, M., Kovatcheva-Datchary, P., Markiewicz, L.H., Stahlman, M., Kozak, L.P., and Backhed, F. (2016). Altered Microbiota Contributes to Reduced Diet-Induced Obesity upon Cold Exposure. Cell Metab *23*, 1216-1223.

Zoetendal, E.G., von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A.D., and de Vos, W.M. (2002). Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. Applied and environmental microbiology *68*, 3401-3407.

Appendix 1: Detailed Discussion of HIV-associated Microbiota Differences

The following text is adapted and updated from (Gootenberg et al., 2017) including data from **Tables 1.5** and **1.6**:

Taxa from four bacterial phyla - Proteobacteria, Fusobacteria, Bacteroidetes, and Firmicutes - have been reported in multiple studies to differ between HIV-positive and HIV-negative individuals. Proteobacteria were more abundant in HIV-infected individuals in 12 of the 35 published studies while others reported no change. The phylum Proteobacteria includes numerous pathogens, such as *Shigella, Salmonella,* and *Helicobacter*. Many specific Proteobacteria genera have been shown to be enriched in HIV-infected individuals and are capable of potentially-pathogenic activities in this context (Zevin et al., 2016): *Pseudomonas* (Gori et al., 2008; Vujkovic-Cvijin et al., 2013) is an opportunistic pathogen that is capable of impairing host mucus production (de Bentzmann and Plesiat, 2011; McGuckin et al., 2011); *Desulfovirbrio* (Lozupone et al., 2013) can produce hydrogen sulfide compounds and inflame the host epithelium (Rey et al., 2013); *Acinetobacter* (Dillon et al., 2014) can produce lipopolysaccharide (LPS)(Peleg et al., 2012) and *in vitro* is able to induce IL-8 production and neutrophil recruitment that could lead to inflammatory tissue damage (Knapp et al., 2006; March et al., 2010); and *Campylobacter* (Mutlu et al., 2014) produces multiple toxins (Walker et al., 1986) that can induce mucosal inflammation (Bereswill et al., 2011). The Proteobacterial family Enterobacteriaceae is associated with inflammation (Guentzel, 1996; Lupp et al., 2007) and was enriched in HIV-infected individuals in 7 of the 35 studies surveyed (most frequently driven by its constituent genus *Escherichia*, which is capable of pathogenic bile acid transformations and degradation of host mucus (McGuckin et al., 2011; Nicholson et al., 2012)). The phylum

Fusobacteria, which is of particular interest due to its previous associations with intestinal inflammation and colorectal cancer (Kostic et al., 2013; Shaw et al., 2016; Sinha et al., 2016), was enriched (most frequently driven by its constituent genus *Fusobacterium*) in HIV-infected individuals in 9 of the 35 previously published studies.

Taxa in the phylum Bacteroidetes, including the families Prevotellaceae, Porphyromonadaceae, Bacteroidaceae and Rikenellaceae, exhibited a more heterogeneous pattern of changes in HIV-infected individuals. The family Prevotellaceae (most frequently driven by its constituent genus *Prevotella*) was enriched in HIV-infected individuals in 12 of 35 studies and depleted in 2 studies. Taxa within the family Prevotellaceae have been associated with inflammation (particularly in the context of autoinflammatory disease (Lukens et al., 2014; Scher et al., 2013)) and activation of gut dendritic cells (Dillon et al., 2016), but a greater abundance of Prevotellaceae is also characteristic of the baseline enteric microbial community of healthy individuals in developing world countries such as Burkina Faso, Venezuela, Malawi, or Papua New Guinea (De Filippo et al., 2010; Gomez et al., 2016; Gorvitovskaia et al., 2016; Lozupone et al., 2013; Martinez et al., 2015; Obregon-Tito et al., 2015; Schnorr et al., 2014; Smits et al., 2017; Wu et al., 2011; Yatsunenko et al., 2012).

Taxa from the family Porphyromonadaceae (most frequently driven by the genera *Barnesiella* or *Odoribacter*) were generally depleted in HIV-infected individuals, exhibiting a decrease in 12 of 35 studies and an increase in 2 studies(Dinh et al., 2015; McHardy et al., 2013). Independent of HIV, Porphyromonadaceae exhibit a diverse and complex array of functions, with both positive (Sinha et al., 2016; Vogtmann et al., 2016;

Zackular et al., 2013) and negative (Zackular et al., 2013) associations with colorectal cancer, and negative associations with *C. difficile* (Milani et al., 2016; Schubert et al., 2014; Schubert et al., 2015), *Salmonella* (Ferreira et al., 2011), vancomycin-resistant *Enterococcus* (Ubeda et al., 2013), and *Citrobacter rodentium* (Bailey et al., 2010; Wlodarska et al., 2011) infection that imply a putative protective role. The family Bacteroidaceae (mostly driven by the abundance of the genus *Bacteroides*) is overall depleted in HIV infection, showing a reduced HIV-associated abundance in 16 of 35 studies(Bender et al., 2016; Dillon et al., 2014; Dillon et al., 2016; Ling et al., 2016; Lozupone et al., 2013; Lozupone et al., 2014; Monaco et al., 2016; Sun et al., 2016; Vázquez-Castellanos et al., 2015; Vujkovic-Cvijin et al., 2013). This family is generally considered to play an anti-inflammatory role (Gauffin Cano et al., 2012; Mazmanian et al., 2008; Round and Mazmanian, 2010; Troy and Kasper, 2010; Vatanen et al., 2016), with the species *Bacteroides fragilis* promoting regulatory T-cell differentiation and IL-10 production via secreted bacterial products (Mazmanian et al., 2008; Round and Mazmanian, 2010; Troy and Kasper, 2010) The family Rikenellaceae (mostly driven by the abundance of the genus *Alistipes*) is also depleted in HIV infection, with decreased abundance in 11 studies and this bile-tolerant family (David et al., 2014) displays protective properties against *C. difficile* infection (Khanna et al., 2016) and a negative association with obesity (Harris et al., 2012) as well as positive associations with both type-1 (Candon et al., 2015) and type-2 (Qin et al., 2012) diabetes mellitus.

Similar to the phylum Bacteroidetes, bacterial families within the phylum Firmicutes were in general reduced in abundance in HIV-infected individuals, though this pattern did not hold true for every family within this phylum. The phylum overall

behaved in this manner, with decreased abundance in 6 of 35 studies, though it was increased in 3 studies. The Firmicutes phylum is quite diverse, but broadly can be characterized as associated with developed world individuals (De Filippo et al., 2010) as well as obesity and increased energy harvest from diet (Ley et al., 2006a; Turnbaugh et al., 2006). Within the Firmicutes, the class Clostridia, which was overall depleted in HIV infection with decreased abundance in 12 of 35 studies and increased abundance in 1 study, is characterized by taxa that often function in anti-inflammatory roles by producing butyrate and other short chain fatty acids (SCFA) (Louis and Flint, 2009; Nicholson et al., 2012) and shifting T-cell differentiation towards regulatory T-cells (Atarashi et al., 2013; Atarashi et al., 2011; Furusawa et al., 2013; Kunisawa and Kiyono, 2011). Within the class Clostridia, the family Lachnospiraceae, which was decreased in abundance in HIV-infected individuals in 12 of 35 studies and increased in 1 study(Dillon et al., 2014; McHardy et al., 2013; Mutlu et al., 2014; Nowak et al., 2015; Sun et al., 2016; Vujkovic-Cvijin et al., 2013), includes members that are commonly found to be uniquely effective metabolizers of complex polysaccharides (Biddle et al., 2013; Flint et al., 2008) and characterized by the production of SCFA such as butyrate (Meehan and Beiko, 2014) and acetate (Rey et al., 2010) that are thought to be anti-inflammatory. Also within the class Clostridia, the family Peptostreptococcaceae varied in its HIV-associated shifts, with 2 of 35 studies (Dubourg et al., 2016; Mutlu et al., 2014)showing a relative decrease in abundance in HIV-infected individuals and 1 study (McHardy et al., 2013)showing a relative increase. Peptostreptococcaceae have been found to function in mostly a pro-inflammatory role, with positive associations with *C. difficile* infection (Milani et al., 2016; Sangster et al., 2016), viral diarrhea (Ma et al.,

2011), intestinal inflammation (Tannock et al., 2012), and the mucosal (Chen et al., 2012) and fecal (Wang et al., 2012) communities of individuals with colorectal cancer. In contrast to the family Peptostreptococcaceae, the Clostridia family Ruminococcaceae was in general decreased in HIV infection, with 16 of 35 studies (Dillon et al., 2014; Dillon et al., 2016; Dubourg et al., 2016; McHardy et al., 2013; Monaco et al., 2016; Mutlu et al., 2014; Sun et al., 2016; Vázquez-Castellanos et al., 2015; Vujkovic-Cvijin et al., 2013; Yu et al., 2014)reporting decreased abundance in HIV-infected individuals and only 2 studies reporting increased abundance. Ruminococcaceae have been associated with both protective and disruptive roles within the gut microbial community, such as the production of anti-inflammatory SCFA (Flint et al., 2008) or the degradation of host mucus and potential pro-inflammatory role in IBD (Png et al., 2010), and functional effects within this family have been found to be highly species dependent (Crost et al., 2013; David et al., 2014; Hsiao et al., 2014; Ze et al., 2012). The bacterial family Erysipelotrichaceae, which is contained within the separate class Erysipelotrichia, was found to be increased in association with HIV infection with a greater abundance demonstrated in 10 of 35 studies. Erysipelotrichaceae are described as adhesive and potentially pathogenic (Turroni et al., 2016), and have been found to be positively associated with obesity (Greiner and Backhed, 2011; Zietak et al., 2016) and luminal microbial communities in colorectal cancer (Chen et al., 2012; Kaakoush, 2015). Interestingly, this family is also found to be enriched in the enteric communities of the Hadza hunter-gatherers of Tanzania (Turroni et al., 2016).