2019

Effects Of Diet And Parasites On The Gut Microbiota Of Diverse Sub-Saharan Africans

Meagan Rubel
University of Pennsylvania, rubel@sas.upenn.edu

Effects Of Diet And Parasites On The Gut Microbiota Of Diverse Sub-Saharan Africans

Abstract

ABSTRACT

EFFECTS OF DIET AND PARASITES ON THE GUT MICROBIOTA OF DIVERSE SUB-SAHARAN AFRICANS

Meagan A. Rubel

Sarah A. Tishkoff

Contemporary African populations possess myriad genetic and phenotypic adaptations to diverse diets, varying climates, and infectious diseases. Most microbiome studies to date have focused on primarily European and Asian populations in urban, industrialized settings. By comparison, relatively little is known about traditional African gut microbiomes, and the range of variation they contain. Many African populations are undergoing substantial changes because of rapid globalization, easier access to hygienic resources and medications, shifts away from traditional lifestyle, and increased exposure to processed diets high in sugars and fats. By characterizing microbiome variation among sub-Saharan African populations using metagenomic sequencing, we can better understand differential response to diseases and environmental factors in producing physiological adaptations. Furthermore, by extending microbiome sampling across a range of traditional African populations in multiple countries, it may be possible to trace subsistence transition with changes in settlement and diet, and interrogate how these are shaped by industrialization. In this dissertation, I describe the gut microbiomes of populations practicing agropastoralism, hunting and gathering, and pastoralism in three African countries: Botswana, Cameroon, and Tanzania. To do this, I used amplicon and shotgun sequencing to characterize microbial genomes and annotate their functions. I combined this microbial data with extensive phenotype and ethnographic data. With this dataset, I detected gut microbial taxa associated with subsistence strategy, sex, geography, and host genetics. I demonstrated that the degree of industrialization in these populations correlated with enrichment of functional pathways involved in the metabolism of xenobiotics and industrial pollutants. Moreover, I found that the gut microbiome has no association with HIV infection, but is highly predictive of multiple gastroenteric parasite infections within Cameroonians. Parasite infection and microbiome composition were, in turn, associated with Th-2 proinflammatory cytokines that are produced during helminthiasis. My research captures microbiota and taxa that are rare or absent from microbiomes of industrialized populations and expands the definition of normal variation within the human gut microbiome. My dissertation identifies multiple factors affecting microbiome composition and works towards generating a more holistic interpretation of the structure and function of human gut microbiota, and their potential associations with human physiology and adaptation.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Anthropology

First Advisor

Sarah A. Tishkoff

EFFECTS OF DIET AND PARASITES ON THE GUT MICROBIOTA OF DIVERSE SUB-SAHARAN AFRICANS

Meagan A. Rubel

A DISSERTATION

in

Anthropology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

_____

Sarah A. Tishkoff, Ph.D.,

Professor of Genetics and Biology

Co-Supervisor of Dissertation

_____

Theodore G. Schurr, Ph.D.,

Professor of Anthropology

Graduate Group Chairperson

_____

Theodore G. Schurr, Professor of Anthropology

Dissertation Committee

Frederic D. Bushman, Ph.D., Professor of Microbiology

Daniel P. Beiting, Ph.D., Assistant Professor of Pathobiology

EFFECTS OF DIET AND PARASITES ON THE GUT MICROBIOTA OF DIVERSE SUB-SAHARAN AFRICANS

For my Dad, who was the first one to show me the wonder of the natural world.

1955-2010

ACKNOWLEDGMENTS

It takes a village, as they say. For an anthropologist studying the gut microbiome with equal measures of parasitology, microbiology, and human genetics thrown in, it has taken a figurative city. First, I owe a debt of gratitude to the many African people who gave time, samples, and energy to this study. Fieldwork was a humbling experience that I am honored to have had. It indelibly changed my research and pedagogy for the better, it humanized my work in a way that no other experience would have, and it let me forge some amazing friendships.

This work would also not be possible without the mentorship of Sarah Tishkoff and Rick Bushman. Sarah allowed me the scientific freedom to pursue an unconventional project in her lab, opened many collaborative doors for my research, and supported me through countless grants, papers, presentations, and international fieldwork. She has been a wellspring of advice throughout my graduate career, and her confidence in my abilities and willingness to probe my assumptions has made me into a more meticulous scientist. Rick Bushman's acumen of technical and experimental knowledge on the microbiome gave me the scientific tools to realize this dissertation. I have greatly appreciated his mentorship, his unwavering commitment to conducting rigorous science, and the energy he's invested in seeing me cross the finish line. Dan Beiting was there in the eleventh hour to offer advice, reagents, computational help, creative solutions, or the pep talk that I needed to overcome obstacles along the way. Tad Schurr paved the path for me to attend the University of Pennsylvania and has chaperoned my scholarship since I started here; his guidance has been essential for my development.

Faculty outside my dissertation committee generously gave time to answer my questions and discuss my research progress. No matter how many projects he was juggling, Kyle Bittinger always made time to offer project advice- he demystified the command line and inspired me with the confidence to begin building my own code. I am indebted to the many talented researchers at the PennCHoP Microbiome Program, especially Ceylan Tanes and Lisa Mattei, for technical assistance throughout this research. My academic sister, Stephanie Schnorr, made anything difficult both endurable and humorous. Our discussions about humans as holobionts have been and continue to be some of the most thought-provoking discussions about human evolution that I've had. Janet Monge and Babette Zemel have been in the wings to champion my cause and talk biological anthropology. Chris Hunter helped me hone my background in parasitology and, along with Omar Harb and Lilach Sheiner, let me TA the Middle Eastern Biology of Parasitism course, which proved to be one of the most personally rewarding endeavors in my academic career thus far. I thank our African collaborators: Thomas Nymabo, Alfred Njamnshi, Charles Fokunang, Gaonyadiwe George Mokone, and Sununguko Wata Mpoloka were instrumental to obtaining permissions and samples in their respective countries, and have been enthusiastic supporters of my research.

I have never met a more intellectually engaging and collaborative group of peers in my life than in graduate school. I am grateful for the camaraderie of my fellow Anthropology graduate students: to my cohort mates, Katherine Culver, Negar Razavi, Utpal Sandesara, and Nick Iacobelli; my fellow bioanthros Volney Friedrich, Sam Larson, Andrea Spence-Aizenberg, Maggie Corley, Raquel Fleskes, Elizabeth Oakley, Clare Super, and

Alexandra Kralick; and Tom Hardy, Kevin Burke, Noam Osband, Kyle Olson, Paul Mitchell, Maddie Boyer, Joanne Baron, Whit Schroeder, Susannah Fishman, Diego Arispe-Bazan, Tiffany Cain, Nooshin Samimi, and Paul Babb- thank you for your brilliance, and your kindness. You all challenged me to situate my work in anthropology, and my research has been all the better for it. The friends I made in the MVP and Genetics programs, including Marco Trazzino, Derek Kelly, Rachel Kember, Matt Hansen, Nick Crawford, Devin Fisher, Shannon Barbour, Eric Mbunwe, Amy Davis, Kelsey Johnson, and Katie Siewert, have been veritable troves of project ideas and morale.

Jaanki Dave was a pleasure to mentor during her summer rotation, and she spent months with me, Elizabeth Loy, and Alexa Avitto in the wet lab trenches, extracting thousands of samples for these projects. The collaborative work we've done together, which goes beyond what's presented in this dissertation, was a genuine pleasure to be a part of. I also want to express my gratitude for all the efforts of Laurie Zimmerman, Mary Kate Hildebrandt, William Beggs, and Sadie Robinson, who helped me navigate the intricacies of Penn and the business of being a graduate student.

It's been a privilege to work with the members of the Bushman and Tishkoff labs, who are the most patient, friendly, and intelligent scientists anyone could ask to work with. For all the endless programming questions you answered, experiment troubleshooting assistance, coffee runs, late-night edits, last-minute lab orders, brainstorming sessions, practice talk attendance, and thought-provoking discussions about science (often over beers), I thank you all from the bottom of my heart. I want to especially recognize Marcia

Holsbach-Beltrame, Louis Taylor, Erik Clarke, Arwa Abbas, Jacob Leiby, Aubrey Bailey, Christel Chehoud, and Jesse Connell for enduring such close proximity to me for so long. I am a much better scientist for having been exposed to all of you, and I hope that the damage incurred on your end of the exchange is minimal.

Samantha Gross, Adam Wilson, Kelly Barse, Jordan Price, Andrew Dodemaide, Matt McAndrews, Terren Proctor, Martha Powers, my Fearless Family, Chris McAdams, Jeff Yost, Dan Moise, Maureen Brown- thank you for reminding me that there is more to life than work, while simultaneously being some of the most fiercely ambitious and caring people that I know. You +1 inspire me.

I won the in-law lottery, and Rita, Tab, Dave, Scotty, Kelly, Ray, and Marie have been cheering me on since Andy first brought me home. So much of the person that I am today is attributable to the love and devotion of my family: Grammy, Uncle Erich, Uncle Billy, Aunt Laurie, Uncle David, Aunt Sue, Tia Yolie, Mandee, Meloni, Ashley, and all of my cousins- you've been with me every step of the way, no matter the distance, and it's meant the world to have your support.

Amber, Mei Mei, and Leela were the best pet therapy anyone could ask for, and they endured more practice talks than anyone else (even if they mostly napped through it).

Most of all, I want to thank Andy and Tara, my husband and my best friend. There wasn't a hard day that went by when I wasn't elevated by your support. Your belief in me has been overwhelming and deeply appreciated. You made Philly home. Thank you, and I love you.

ABSTRACT

EFFECTS OF DIET AND PARASITES ON THE GUT MICROBIOTA OF DIVERSE SUB-SAHARAN AFRICANS

Meagan A. Rubel

Sarah A. Tishkoff

Contemporary African populations possess myriad genetic and phenotypic adaptations to diverse diets, varying climates, and infectious diseases. Most microbiome studies to date have focused on primarily European and Asian populations in urban, industrialized settings. By comparison, relatively little is known about traditional African gut microbiomes, and the range of variation they contain. Many African populations are undergoing substantial changes because of rapid globalization, easier access to hygienic resources and medications, shifts away from traditional lifestyle, and increased exposure to processed diets high in sugars and fats. By characterizing microbiome variation among sub-Saharan African populations using metagenomic sequencing, we can better understand differential response to diseases and environmental factors in producing physiological adaptations. Furthermore, by extending microbiome sampling across a range of traditional African populations in multiple countries, it may be possible to trace subsistence transition with changes in settlement and diet, and interrogate how these are shaped by industrialization. In this dissertation, I describe the gut microbiomes of populations practicing agropastoralism, hunting and gathering, and pastoralism in three African countries: Botswana, Cameroon, and Tanzania. To do this, I used amplicon and shotgun sequencing to characterize microbial genomes and annotate their functions. I

combined this microbial data with extensive phenotype and ethnographic data. With this dataset, I detected gut microbial taxa associated with subsistence strategy, sex, geography, and host genetics. I demonstrated that the degree of industrialization in these populations correlated with enrichment of functional pathways involved in the metabolism of xenobiotics and industrial pollutants. Moreover, I found that the gut microbiome has no association with HIV infection, but is highly predictive of multiple gastroenteric parasite infections within Cameroonians. Parasite infection and microbiome composition were, in turn, associated with Th-2 proinflammatory cytokines that are produced during helminthiasis. My research captures microbiota and taxa that are rare or absent from microbiomes of industrialized populations and expands the definition of normal variation within the human gut microbiome. My dissertation identifies multiple factors affecting microbiome composition and works towards generating a more holistic interpretation of the structure and function of human gut microbiota, and their potential associations with human physiology and adaptation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF SUPPLEMENTAL FILES

Chapter 1. Introduction

This dissertation expands knowledge on the impact of subsistence, geography, pathogen infection, and ancestry on gut microbiome composition in diverse, rural African populations. In Chapter 2, I provide a broad background about human evolutionary history in Africa, with a focus on *Homo sapiens*, through the lens of '-omics research. I then provide background on the gut microbiome, including its association with health and disease states in humans and a description of the tools that are used to interrogate the gut microbiome. Finally, this chapter concludes with a description of substantial health problems in Africa, with an emphasis on infectious disease, and their known correlates with gut microbiome composition and function.

In Chapters 3 and 4, I describe the characterization of the gut microbiomes of diverse sub-Saharan African populations, focusing on the roles of subsistence, geography, genetics, and human pathogens. My aim is to provide insights into the co-evolutionary relationships between people, their enteric microbiota, pathogens, and the external environment (diet, geography).

Specifically, I test the following hypotheses:

1) Rural African gut microbiome composition will be adapted to different subsistence groups, diets, local geography, sex, and host genotype (Chapters 3 and 4).
2) Consequent to their level of industrialization, agropastoralist and pastoralist populations will have microbial genes that are enriched for functions in the

degradation of chemicals, including industrial pollutants and antibiotics, as compared to hunter-gatherer populations (Chapters 3 and 4).

3) Human gut microbiome composition will have correlations with pathogen presence and frequency, including blood/fecal parasite infection and HIV. Infected individuals will possess microbiota with pro- or anti-inflammatory functions associated with their physiological response to pathogen infection, which can be further evidenced by cytokine and other biomarker responses in the host (Chapter 4).

In Chapter 3, I describe the gut microbiota from six rural populations with agropastoral, pastoral, and hunter-gatherer subsistence from Botswana and Tanzania and compare them with an urban U.S. population. I demonstrate that the Tanzanians harbor the most individual diversity in their gut microbiomes, and that hunter-gatherer microbiomes are phylogenetically distinct from the microbiomes of pastoralist and agropastoralist groups. Pastoralists, whose diets are enriched in meat, blood, and dairy, do not have significantly different microbiomes from agropastoralists, indicating the potential convergence in microbiota types required to catabolize their different diets. Although genetic relatedness is correlated with bacterial composition for two African populations, I find that geographic proximity varies more significantly with bacterial abundance. Furthermore, I find evidence for sex-specific taxa in two populations, the Maasai and Hadza, who are known to practice sexual-specific division of labor. Bantu-speaking agropastoralists from Botswana have gut bacteria very similar to U.S. individuals, with imputed metagenomic

content that is enriched for genes found in metabolic pathways for industrial pollutant degradation.

In Chapter 4, I test for associations between the gut microbiomes of traditional African populations with their pathogens Most studies tend to pursue helminth-microbiota research in one of three ways: They either classify variation between case and control groups of helminth infected individuals, to treat with specific bacteria to test for inhibitory effects on parasite survival or map pro- and anti-inflammatory physiological response from bacterial changes associated with parasite infection. Conflicting results in the microbiota associated with parasite positive and negative individuals thus necessitate reproducible results on large, global cohorts to identify shifts in taxa related to pathogens. I quantify parasite copy genome number in Cameroonians for seven parasites: *Ascaris lumbricoides,* Pan-*Cryptosporidum spp., Entamoeba histolytica, Giardia lamblia, Necator americanus, Strongyloides stercoralis,* and *Trichuris trichiura,* and I test for HIV and cytokine levels. I use extensive phenotype and ethnographic data to create case control matches for different types and combinations of parasites in Cameroonians representing three subsistence groups (hunter-gatherers, pastoralists, and agropastoralists). These data are compared with the microbiomes of healthy individuals from the industrialized U.S. The results indicate that parasite infection in Cameroonian populations was common, especially in hunter-gatherers, and that bacterial diversity was positively correlated with frequency of parasite infections. Microbiome composition could be used to predict parasite infections with a significantly co-occurring group, *A. lumbricoides, N. americanus, T. trichiura,* and *S. stercoralis,* dubbed the “ANTS” group,

with high accuracy. I also found multiple infections of pathogenic and commensal *Entamoeba* species that were not detected in traditional qPCR and note that microbiome composition can be used to predict *Entamoeba* infection, although with less accuracy than ANTS infection can be predicted. There was no association with either HIV infection or the lactose persistence phenotype with the gut microbiome, although the pastoralists had an enrichment of bacteria capable of catabolizing the products of lactose metabolism (galactose). Finally, supervised machine-learning models using microbiome abundance to separately predict IL-5 cytokine levels and different categories of ANTS produced overlapping results. Some microbial taxa were highly predictive of both IL-5 and ANTS status, and warrant further investigation for their role in promoting or decreasing helminth-associated physiological morbidity.

Through the study of contemporary groups with different, traditional subsistence practices, such as those described here, ethnographic analogies can be developed for how ancestral populations lived. The characterization of gut microbiota from these populations- which, for some groups, marks the first time their gut microbiota has been sequenced- may inform on how humans have adapted to shifting environments and selective climatic, dietary, and infectious pressures in sub-Saharan Africa. Here, we find evidence for associations between the gut microbiota with sex, genetics, geography, subsistence, multiple types of infection, and cytokines. Many of these specific associations have not been described in the human gut microbiome before. Given the rapid rate of globalization in many parts of sub-Saharan Africa, and the microbial changes consequent to the processes of industrialization, this dissertation serves as a

cross-sectional study on the changing landscape of human gut microbiota in diverse, rural populations.

Chapter 2. African evolutionary history, health, and the gut microbiome

Section 2.1 of this chapter has been modified from a previously published article:

> Holsbach-Beltrame[†], M., Rubel[†], M.A. & Tishkoff, S.A. (2016). Inferences of African evolutionary history from genomic data. *Curr. Opin. Genet. Dev*. doi:10.1016/j.gde.2016.10.002
>
> [†] Contributed equally

2.1 Inferences of African evolutionary history from genomic data

Using '-omics' to understand human origins in sub-Saharan Africa

Archaeological and genetic studies indicate that Africa is the origin of anatomically modern humans (AMH) within the past 300 thousand years (ky), the source of the worldwide range expansion of modern humans in the past 100 ky, and contains the greatest levels of human genetic variation on a global scale (Campbell & Tishkoff, 2008, 2010; Hublin et al., 2017; Reyes-Centeno, 2016; The 1000 Genomes Project Consortium, 2015). Africa contains substantial linguistic and cultural diversity, with populations that traditionally eat diverse diets and practice various subsistence methods (e.g. agriculturalists, hunter-gatherers, and pastoralists) (Beltrame, Rubel, & Tishkoff, 2016; Campbell & Tishkoff, 2008, 2010; Tishkoff et al., 2009). Geographically and ethnically diverse African populations have been exposed to distinct selective pressures through dramatically different environments, climates, diet, and pathogen exposures. Some of these genetic adaptations are known to play roles in diet (e.g. amylase copy number variation, lactase persistence, bitter taste perception) and diseases (e.g. hemoglobinopathies) (Campbell et al., 2012; Campbell & Tishkoff, 2008; Perry et al.,

2007; Scheinfeldt et al., 2012; Tishkoff, Reed, et al., 2007). Despite the wide spectrum of phenotypic and genotypic diversity (M. Jakobsson et al., 2008; Ramachandran et al., 2005; Schlebusch et al., 2012; Tishkoff et al., 2009), sub-Saharan Africans (SSA) remain underrepresented and understudied for analyses of adaptive traits and genetic variation. This research deficit, in turn, has produced gaps in the understanding of African disease susceptibility and population history.

As the development and cost of sequencing technologies continues to decrease and more sophisticated computational models are developed, researchers are increasingly able to decipher sub-Saharan genetic and biological structure and function to understand human evolutionary history on a finer and more cohesive scale than ever before. This section summarizes how omics-technologies (specifically genomics, methylomics and microbiomics) and analyses of SSA shed light on (1) AMH origins and archaic human introgression (gene flow between species via repeated hybrid backcrossing with parent species), (2) patterns of diversity and substructure in contemporary SSA populations, and (3) SSA genetic adaptations and disease.

As AMH left Africa and migrated through Eurasia, they encountered now-extinct archaic populations (i.e., Neanderthals and Denisovans), with gene flow events occurring between AMH and archaic populations at least once and potentially multiple times from 37 to 96 kya (B. Y. Kim & Lohmueller, 2015; Mafessoni, 2019; Nielsen et al., 2017; Sankararaman, Patterson, Li, Pääbo, & Reich, 2012; Vernot & Akey, 2014, 2015; Wolf & Akey, 2018). As a result, ~2–6% of the genomes of non-African individuals derive from

Neanderthals (Green et al., 2010) and/or Denisovans (Reich et al., 2010). However, less is known about archaic admixture in early modern human groups residing in SSA. AMH fossils from the Jebel Irhoud (Hublin et al., 2017), Herto (Clark et al., 2003; White et al., 2003) and Omo Kibish (Aubert et al., 2012; McDougall, Brown, & Fleagle, 2005) sites date to ~300 kya, 160 kya, and 195 kya, respectively (Figure 2.1). These AMH fossils overlap with other Pleistocene fossils from transitional *Homo* genera that show a range of archaic and modern traits. Such transitional fossils support the coexistence of AMH with other, morphologically differentiated forms of archaic hominins until ~35 kya (Bräuer, 2008; Harvati et al., 2011; Rightmire, 2009).

Buttressing fossil evidence (Harvati et al., 2011; Rightmire, 2009), various genetic studies have proposed that anywhere from 2-5% of African genomes were contributed from a now-extinct taxon of the genus *Homo* (Hammer, Woerner, Mendez, Watkins, & Wall, 2011; Hsieh, Woerner, et al., 2016; Lachance et al., 2012; Plagnol & Wall, 2006; Stringer, Harvati, Allsworth-Jones, Grün, & Folorunso, 2010; Wall, Lohmueller, & Plagnol, 2009; Xu et al., 2017), and that archaic lineages may have persisted as late as 25-10 kya (Hammer et al., 2011; Harvati et al., 2011). While a direct comparison between a full archaic African reference sequence and modern African genomes would provide unambiguous evidence for interbreeding, DNA from ancient samples remains challenging to obtain as fossils rapidly decay in the tropical environments found throughout much of SSA. Thus, indirect approaches to identify and model ancient admixture have been developed to discern signals of introgression within SSA populations (Hammer et al., 2011, p. 201; Lachance et al., 2012; Plagnol & Wall, 2006;

Wall et al., 2013). Hammer and colleagues (2011) examined multilocus DNA sequence polymorphism data from 61 non-coding regions in one West African population (Mandinka), a Central African rainforest inhabiting hunter-gatherer (CAHG) population (Biaka) and southern African San populations and detected signals of archaic introgression in the Biaka and San, but not the Mandinka. Approximate-likelihood statistical models, with and without gene flow from archaic populations, were used to infer that ~2% of genetic material found in contemporary African populations was introgressed ~35 kya from a group of unspecified archaic hominins that separated from the ancestors of AMH ~700 kya during the Lower-Middle Pleistocene (Hammer et al., 2011).

To explore these observations, Lachance et al., 2012 conducted high-coverage whole-genome sequencing on fifteen hunter-gatherers from three ethnic groups, including CAHG from Cameroon characterized by a short-statured 'pygmy' phenotype, Hadza hunter-gatherers from Tanzania, and the Sandawe from Tanzania, who until recently practiced a hunting and gathering lifestyle. Analysis of these populations revealed overlapping introgressed genetic regions in all three populations from at least one archaic population. Using the putative introgressed haplotypes, Lachance et al. (2012) noted a median time to most recent common ancestor (TMRCA) of 1.2–1.3 million years ago (mya), which is similar to the TMRCA for introgressed haplotypes in Europeans (1.1–1.2 mya) and indicates a similar timeframe of divergence of archaic populations in Africa and Europe (Neanderthals) from AMH.

Similarly, Hsieh and colleagues (2016) used whole genomic sequences from seven hunter-gatherers (three Baka, four Biaka) to statistically model isolation and gene flow events. Their results revealed 265 candidate introgressed loci with a median TMRCA of 1.08 mya, which was compatible with prior ranges (Lachance et al., 2012). Furthermore, Hsieh et al. (2016) found support for at least one, and possibly recurrent, admixture event(s) that favored a model of low level, frequent interbreeding between archaic and modern humans in Africa.

Future studies of genomic introgression of modern and archaic populations in Africa would benefit from the inclusion of more ancient human genomes from SSA fossils and larger samples of modern human genomes from diverse SSA populations. To date, three studies have conducted aDNA research on SSA populations. The first sequenced the genome of a 4,000-year old AMH fossil, nicknamed “Mota,” from Ethiopia (Llorente et al., 2015), the second conducted genome-wide analysis on 16 African individuals who lived ~8 kya (Skoglund et al., 2017), and the third focused on the origins of pastoralism in ~5,000-year old East African samples (Prendergast et al., 2019). The latter study yielded evidence of genomic adaptations to bitter taste perception and growth in modern and ancient sub-Saharan Africans, indicated that hunting and gathering populations related to the South African San were once widespread in eastern Africa, and showed that a divergent lineage contributed to the genomes of western Africans. Although informative for reconstructing recent events, analyses of additional ancient fossils will be necessary to infer introgression events. Screening for introgressed variants that are adaptive, as has been done with Neanderthal variants in extant European and East Asian

populations (Mathieson et al., 2015; Sankararaman et al., 2012; Vernot & Akey, 2015), may reveal whether archaic alleles played a role in adaptation in SSA populations.

Patterns of diversity and substructure in sub-Saharan Africans

The patterns of genetic diversity in SSA populations reflect their complex demographic history of short-range and long-range migration events, subdivision and admixture, as well as local adaptation to a diverse array of environments. The genetic diversity of African populations parallels their linguistic diversity, with more than 2,000 languages being spoken across the continent (Ehret, 2000). The majority of these languages belong to four major families: Niger-Kordofanian, which includes Bantu languages and predominates throughout most of SSA; Afroasiatic, which is spoken in Saharan, northeastern and eastern Africa; Nilo-Saharan, which is spoken in portions of Saharan, eastern, and northeastern Africa; and Khoesan, a family of languages denoted by their click consonants and spoken by the San in southern Africa and the Hadza and Sandawe in eastern Africa (Boyeldieu et al., 2008; Heine & Nurse, 2000; Sands, 1998; Tishkoff et al., 2009) (Figure 2.1).

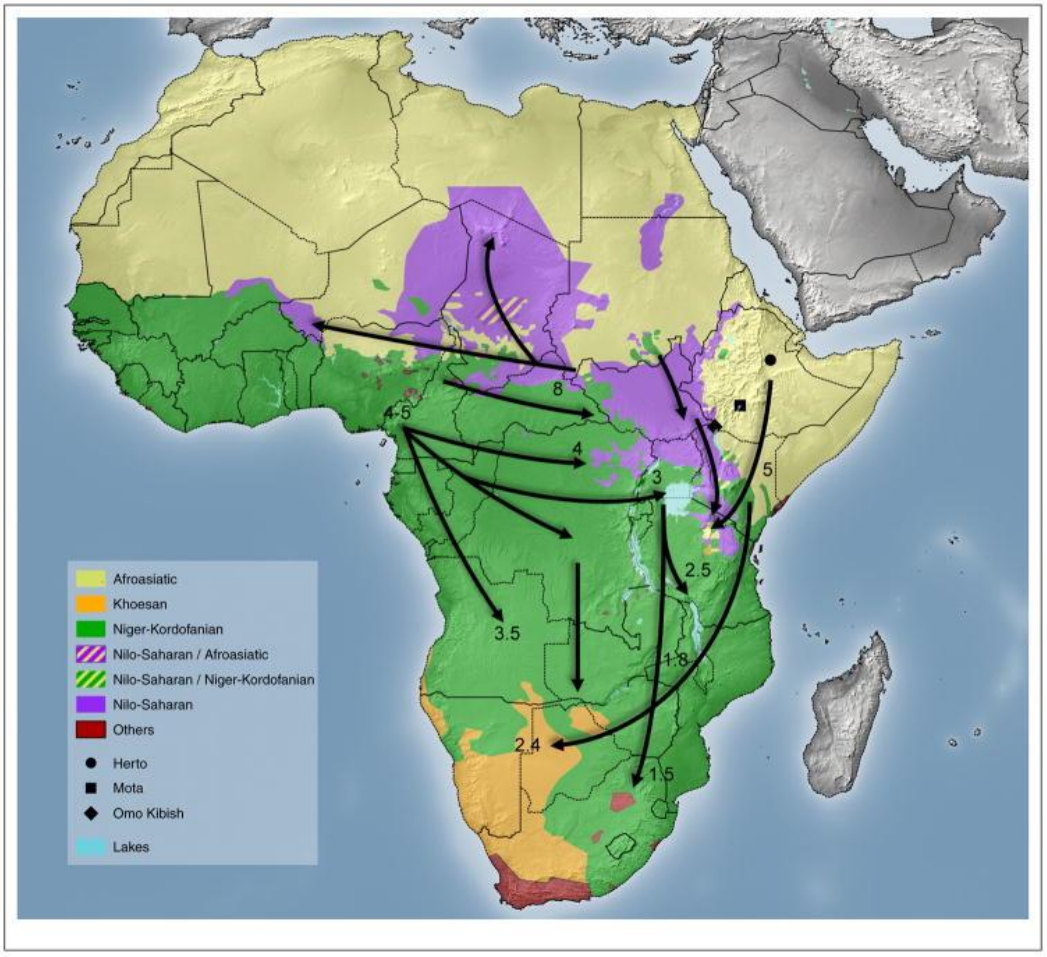


Figure 2-1. Map of Africa showing the distribution of the major language families, the location of hominid remains discussed in the text, and major migration routes of AMH through the continent within the past 10 kya. Numbers on the map denote how old different migratory events are in thousands of years.The expansion of Bantu-speaking people (referred to as the "Bantu expansion") started around 5–4 kya from the Cameroon/Nigeria border, initially from west to east, either north or south of the rainforest, to the Great Lakes of Uganda by around 3 kya and then from east to south in the last 2.5 ky, rapidly expanding into central and southern Africa, reaching Mozambique ~1.8 kya and South Africa ~1.5 kya. Another dispersion occurred ~3.5 kya from Cameroon, moving south to Angola. The spread of pastoralism into sub-Saharan Africa occurred around 4.5 kya (for a review see Pedro Soares, Rito, Pereira, & Richards, 2016). Afroasiatic-speaking agropastoralist populations migrated from Ethiopia into Kenya and Tanzania within the past 5 ky (Patin et al., 2009). After admixing with Bantu groups, pastoralist populations from eastern Africa migrated through Tanzania to southern Africa around 2.4 kya (Gurdasani et al., 2015; Pickrell et al., 2014). Migrations through the Sahel occurred bidirectionally between east and west Africa in the past 8 ky (Gibbs et al., 2003; Hirbo, Ranciaro, & Tishkoff, 2012). There are many fossil sites of archaic and early AMH populations in Africa. A handful of sites important to discussions on human origins and archaic introgression are listed

here. Herto: The well-preserved cranial remains of three individuals, dated to ~160 kya, were recovered at Herto Bouri, near the Middle Awash of Ethiopia. The crania are robust, consistent with other large archaic subspecies of *Homo* fossils found in Africa, but have endocranial morphology and facial features typical of modern *Homo sapiens;* thus they are thought to represent an extinct subspecies of *Homo sapiens*. Mota: The first ancient DNA analysis from Africa came from a ~4.5 ky old fossil found in the Mota Cave in the Ethiopian highlands (Llorente et al., 2015). This fossil set the timing of reverse gene flow from Eurasia into Eastern Africa AMH ~1.5 ky earlier than prior estimates. Omo Kibish: Two crania, named Omo 1 and Omo 2, as well as post-cranial skeletal elements were found near the Omo River in Ethiopia. Dated to 195 kya, they possess a cranial vault height similar to AMH, situating them as early members of the Homo sapiens lineage. Of the two crania, Omo 1 has stronger morphological similarity to modern humans.

The International HapMap Project (Gibbs et al., 2003) and the Human Genome Diversity Project (HGDP) (Cann et al., 2002) represent some of the earliest large-scale efforts by research consortiums to catalog human genetic diversity. More recently, the international "1000 Genomes Project" (The 1000 Genomes Project Consortium, 2015) has included an expanded range of populations to compile a global reference of human genetic variation. Although twelve African populations are represented between these projects, with overlap of Nigerian Yoruba in all three, most of the African populations are of recent Niger-Kordofanian ancestry and do not reflect the range of diversity present in Africa. To rectify this problem, the African Genome Variation Project (AGVP) is currently conducting dense genotyping and whole-genome sequencing across individuals belonging to ten language subgroups in sub-Saharan Africa; however, their low coverage genomic sequencing (4x coverage) risks misclassifying rare variants (Gurdasani et al., 2015). In addition, statistically imputing genetic variants from low coverage genome data, and from genetically sub-structured African populations, can be challenging. The largest high coverage whole genome sequencing data set, representing 44 indigenous African populations, identified millions of novel genetic variants and regions contributing to local adaptation among populations (Fan et al., 2019).

There have been efforts to expand studies of genome-wide genetic markers to include broader groups of sub-Saharan African populations. Analysis of genetic substructure based on genome-wide microsatellite and insertion/deletion markers in 121 geographically diverse African populations identified fourteen ancestral population clusters that were correlated with self-described ethnicity and shared cultural or linguistic features (Tishkoff et al., 2009). The results of this study and others suggested that African populations have maintained a large and subdivided population structure throughout much of their evolutionary history (reviewed in Hirbo, Ranciaro, and Tishkoff 2012). More recent genome wide SNP genotyping studies have largely supported these observations. Principal components analysis (PCA) indicates that, generally, SSA populations cluster based on their geographic distribution (Bryc et al., 2010; Busby et al., 2016; Montinaro, Busby, Gonzalez-Santos, & Oosthuitzen, 2016; Tishkoff et al., 2009; Uren et al., 2016) with some exceptions. For example, the CAHG cluster near the southern African San, suggesting that they may once have ancient shared common ancestry (Tishkoff et al., 2009; Veeramah et al., 2012). There has also been extensive admixture in African populations (Hellenthal et al., 2014; Tishkoff et al., 2009). Much of this admixture is driven by the Bantu expansion, which marked a series of major human migrations in the last 4000 years from west Africa (Soares et al., 2016) (Figure 2.1). The Bantu migration began around the Nigeria/Cameroon highlands and moved to the east and then to the south of the continent (reviewed in Hirbo, Ranciaro, and Tishkoff 2012) resulting in the expansion of agricultural practices, and the displacement, replacement, or admixture with other populations in many regions of SSA (Busby et al., 2016; Li, Schlebusch, & Jakobsson, 2014; Tishkoff et al., 2009).

Several studies of SSA populations have included analysis of African groups that currently, or until recently, have practiced hunting and gathering subsistence. Autosomal sequencing and SNP array data indicate that the genetic lineages of Khoesan-speaking populations of southern Africa are basal to other populations, with an estimated divergence time around 157-108 kya (Fan et al., 2019; Lachance et al., 2012; Pickrell et al., 2012; Schlebusch et al., 2012; Veeramah et al., 2012; Veeramah & Hammer, 2014). Furthermore, genome-wide SNP array studies indicate extensive population substructure in southern Africa (Montinaro et al., 2016; Pickrell et al., 2012; Uren et al., 2016). A recent analysis of genome-wide SNPs from 21 southern African populations indicate that this structure was the byproduct of geographical and ecological barriers around the Kalahari Basin (Uren et al., 2016). Pickrell et al. (2012) inferred that the southern Khoesan-speaking San split into two geographic groups in the northwestern and southeastern Kalahari ~30 kya. Challenging this observation, a recent study of whole-genome SNP data from 46 African populations found evidence for at least three genetically divergent Khoesan groups (northern, central, and southern Kalahari), and dated their divergence to ~33 kya (Montinaro et al., 2016). CAHG populations diverged from the San ~48 - 60 kya (Patin et al., 2009; Veeramah et al., 2012) and the Hadza and Sandawe diverged from other populations > 15 kya (Tishkoff, Gonder, et al., 2007).

Adaptation in sub-Saharan Africa

African populations have adapted to a broad range of environments and diets and are likely to have regional- or population-specific adaptive traits. Many statistical approaches have been developed to detect recent selective sweeps (i.e. iHS and EHH) or local

adaptation (i.e. F_{ST}, LSBL, PSB, XPCLR) (reviewed in Scheinfeldt & Tishkoff, 2013) and Fan et al. (2016). If these targets of selection fall within a genomic region associated with a putative function, the adaptive phenotype can potentially be inferred. Functional significance of an adaptive variant can be determined by *in vitro* studies of gene expression, in the case of regulatory variants, *in vivo* analyses in model organisms (e.g., genetic variation at *MFSD12* was associated with reduced pheomelanin in mice (*Mfsd12),* and corresponded to darker skin pigmentation in Africans (Crawford et al., 2017), or by genotype/phenotype association studies. Some prominent examples which integrate genome-wide scans of selection with genotype/phenotype association studies or other functional assays in Africans include bitter taste perception (Campbell et al., 2012), adaptation to diverse diets (Breton et al., 2014; Perry et al., 2007; Ranciaro et al., 2014; Tishkoff, Gonder, et al., 2007), high-altitude adaptation (Alkorta-Aranburu et al., 2012; Scheinfeldt et al., 2012), short stature in CAHG (Jarvis et al., 2012; Lachance et al., 2012; Perry et al., 2014), skin pigmentation (Crawford et al., 2017), and immune response (Genovese et al., 2010; Jallow et al., 2009; Ko et al., 2011; Lachance, 2010; Schlebusch et al., 2012; Timmann et al., 2012; Tishkoff et al., 2001), a few of which are discussed here in greater detail (Figure 2.2).

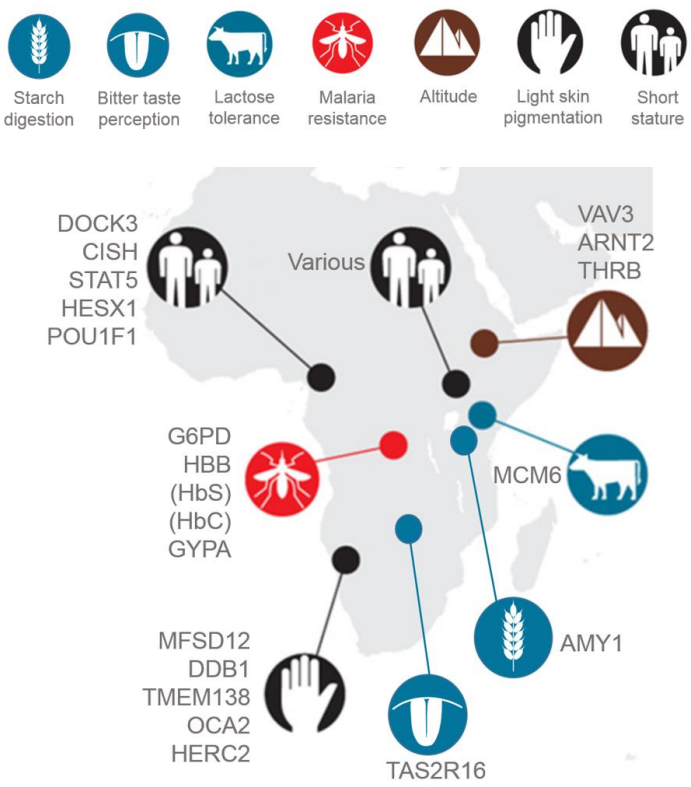


Figure 2-2. Distribution of locally adaptive traits within Africa. Adaptation to diverse environments during human evolution has resulted in phenotypes that are at the extremes of the global phenotype distribution. Integrated scans of natural selection and GWAS have been used to identify genomic regions that influence trait variation. Within Africa, these include genes, allelic variants, and copy number variants associated with malaria resistance (G6PD, HBB, HbS, HbC, GYPA), short stature (DOCK3, CISH, STAT5, HES1, POU1F1, others), high altitude adaptation (VAV3, ARNT2, THRB), lactase persistence (MCM6), skin pigmentation (MFSD12, DDB1, TMEM138, OCA2, HERC2), bitter taste perception (TAS2R16), and starch digestion (AMY1). This figure has been modified from a previously published version in Rubel, M.A., Cuadra A., Illustration for Tishkoff (2015). Strength in small numbers. doi: 10.1126/science.aad0584.

Genome-wide scans for selection have revealed extended haplotype homozygosity on chromosome 2 in the region containing lactase persistence (LP) associated alleles in the genomes of Europeans and Africans (Bersaglieri et al., 2004; Voight, Kudaravalli, Wen, & Pritchard, 2006). LP confers the ability to digest lactose (the sugar present in milk) during adulthood, giving a selective advantage to those individuals who can consume

dairy products as a source of nutrition and hydration. Four alleles associated with LP have been demonstrated to be functional, and they are located in a regulatory region in intron 13 of the *MCM6* gene, an enhancer of the lactase gene (*LCT*) (Hassan et al., 2016; Labrie et al., 2016). The variant *T-13910* is common in Europe, but is also present at low levels in the Middle East and in Central and Western Africa, where it was introduced through migration (Bersaglieri et al., 2004; Breton et al., 2014). In eastern/northeastern Africa, three functionally relevant genetic variations have been identified: *C-14010*, *G-13915* and *G-13907* (Ingram et al., 2007; Macholdt, Slatkin, Pakendorf, & Stoneking, 2015; Ranciaro et al., 2014; Tishkoff, Reed, et al., 2007). The presence of African and European LP-associated variants on different haplotype backgrounds shows that the LP trait has evolved independently multiple times, a classic example of convergent evolution (Tishkoff, Reed, et al., 2007).

Genomic scans for selection have also been applied to complex traits that are influenced by multiple genetic variants and environment, such as stature. For example, the short stature phenotype of CAHG populations has been proposed to be an adaptive phenotype for a tropical forest environment (Hsieh, Veeramah, et al., 2016; Jarvis et al., 2012; Lachance et al., 2012; Migliano, Romero, Leavesley, & Pagani, 2013; Perry et al., 2014; Pickrell et al., 2009). Jarvis et al. (2012) genotyped CAHG from Cameroon and neighboring Bantu groups using a genome-wide SNP array and identified signatures of positive selection in CAHG for regions enriched in genes involved with immune response, reproduction, thyroid function, and body size. A subset of these loci (e.g., *DOCK3* and *CISH*) was previously associated with stature. Lachance et al. (2012) looked

for differentiated regions of the genome in CAHG using high coverage whole genome sequence data. Genes in the top differentiated regions include *TRHR,* which is expressed on the surface of anterior pituitary cells and has an important role in thyroid function, and *HESX1,* which encodes a homeobox-containing transcriptional repressor that plays a critical role in development of the anterior pituitary, the site of growth hormone synthesis and secretion. Other targets of selection were near genes involved in immunity, metabolism, fertility, and olfaction. Perry et al. (2014) genotyped Batwa CAHG and neighboring Bakiga agriculturalists from Uganda using a genome-wide SNP array. Consistent with results from Jarvis et al. (2012) and Lachance et al. (2012), Perry et al. (2014) found several potential signals of selection, including a 15 Mb region on chromosome 3. They identified sixteen genomic regions marginally associated with short stature. These results did not replicate those from a group of western CAHG (Baka), leading Perry et al. (2014) to suggest that the "pygmy" phenotype may be an example of convergent evolution in western and eastern CAHG groups. The identification of loci associated with short stature in CAHG, and the question of convergent evolution in western and eastern CAHG populations remains an ongoing area of investigation.

Two recent studies (Fagny et al., 2015; Gopalan et al., 2017) investigated patterns of methylation across the genome (e.g. "methylome") of CAHG and Bantu groups. These studies revealed differentially methylated loci near genes that play a role in immunity, glucose and lipid metabolism, fatty acid, bone growth, and stature. They also identified genetic variants influencing differential methylation (known as methylation quantitative loci, or meQTLs) in genes involved in insulin metabolism, bone-mineral density, and

height. Gopalan et al. (2016) found patterns of methylation in African hunter-gatherers (Baka CAHG and San) that could be useful for inferring age, particularly in geographic regions where age is often unknown. Taken together, these epigenomic studies highlight how genetic variation and the environment can influence methylation patterns, which in turn can influence adaptive trait variation.

Lastly, gut microbiome studies in SSA are informative for distinguishing how humans and bacteria have co-evolved. These studies have indicated that gut microbiomes of hunter-gatherers are generally distinct from western populations and are abundant in microbiota that may be particular to their traditional diets (Ayeni et al., 2018; De Filippo et al., 2010; Gomez et al., 2016; Hansen et al., 2019; Morton et al., 2015; Rampelli et al., 2015; Schnorr et al., 2014; Smits et al., 2017; Yatsunenko et al., 2012). A detailed discussion on the gut microbiomes of non-industrialized populations is given in section 1.3. Broadly, while these studies indicate that subsistence practice and pathogen infection shape the microbial composition of the gut, an expansion of this research to include more studies of ancient human microbial communities could elucidate differences between modern and ancient microbe structure and function, and provide evidence for adaptation to changing environmental pressures and dietary shifts during long term human evolution.

Advances in ‘-omics technologies are generating an unprecedented level of knowledge about human genomes and physiology, disease susceptibility, and evolutionary history, but their application to SSA populations is still disproportionately small. The analysis of large sample sizes in diverse African populations is becoming increasingly possible and

financially feasible with next-generation '-omics approaches. Integrating multiple methods to detect local adaptation to diverse environments, such as GWAS and genome-wide scans for selection, takes advantage of the variable environments, demography, and population structure present in SSA. There is a substantiated need to develop better demographic models that provide accurate estimates of divergence times, migration events, and fluctuations in population size in African populations, as well as the ability to discern these genome-wide demographic effects from locus-specific selection. Combining these approaches, together with novel integrative genomics studies that incorporate data from proteomics, metabolomics, transcriptomics, epigenomics and microbiomics, will shed light on the complex demographic and adaptive history of sub-Saharan African populations.

2.2 Metagenomics and the Multifactorial Gut Microbiome

Microbiota of the distal gastrointestinal tract in humans

The human microbiome is composed of bacterial, viral, archaeal, protist, and unicellular fungal microbes found on and within human hosts. There is an estimated ~1:1 ratio of human to bacterial cells in the human body (Sender, Fuchs, & Milo, 2016), and among human microbiota sites, the gastrointestinal (GI) tract contains the most bacterial diversity (Ley, Peterson, & Gordon, 2006; O'Hara & Shanahan, 2006; Wilson & Nicholson, 2009). The gut microbiome consists of microbes which interact with one another as well as their hosts (Clemente, Ursell, Parfrey, & Knight, 2012; Qin et al., 2010). Consequently, gut microbiome composition has significant implications for

human disease and health with regard to nutrition, digestion, metabolism, and immune response (Kelly, 2010; Walter & Ley, 2011; Wilson & Nicholson, 2009).

The GI tract is generally divided into the upper GI tract (esophagus, stomach, and duodenum), the lower GI tract (small intestine, large intestine), and their accessory organs. Due to its restricted accessibility, the microbiomes of the human stomach and small intestine have been less intensively investigated than the large intestine microbiome (Marchesi, 2014). The large intestine resorbs ions and water from gastric juices and partially digested food (known as "chyme"), which is processed into fecal matter. The density of resident microbiota in the large intestine is greater than any other human microbiome (10^{12} cells per gram of intestinal content, 30% of large intestine volume), and they are responsible for processing nutrients, mucus, digestive enzymes, and shed epithelial cells (Marchesi, 2014).

High-throughput sequencing of the taxonomically informative bacterial 16S ribosomal RNA gene (16S rRNA) has revealed that there could be at least 1,800 genera and between 15,000 to 36,000 species of bacteria in the large intestine (Frank et al., 2007). Obligate anaerobes dominate the large intestine, with two phyla, *Firmicutes* and *Bacteroidetes,* comprising more than 80% of all phylotypes, or groups of DNA sequences sharing similar gene markers (Frank et al., 2007). The *Firmicutes* are comprised of the *Clostridium coccoides* and *C. leptum* families (among others), while *Bacteroidetes* is affiliated with genera including *Bacteroides*. Other bacterial taxa occupy the lower intestine in comparatively smaller numbers (Arumugam et al., 2011). The microbiota of the large intestine produce short chain fatty acids (SCFA) (butyrate, acetate, and

propionate) from carbohydrate and mucin fermentation end-products, biosynthesize vitamin K, convert equol (an isoflavandiol estrogen metabolized from daidzein), and are involved in various drug and nutrient metabolism activities (Marchesi, 2014; Wilson & Nicholson, 2009).

Archaea are present in smaller numbers than bacteria in the microbiome (Matarazzo et al., 2012). Archaea are dominated by *Methanobrevibacter smithii*, which degrades complex plant polysaccharides and produces methane as a byproduct of fermentation (Stewart, Chadwick, & Murray, 2006). Much of what is known about archaea in the gut is focused on the abundant *M. smithii*, and to a lesser extent other methane producing archaea, or "Methanogens." Research on methanogens indicates that they increase in humans over the first years of life, have stable abundances during early adult years, and display high levels of diversity in elderly adults (Dridi, Fardeau, Ollivier, Raoult, & Drancourt, 2011; Lewis & Cochrane, 2007; Mihajlovski, Doré, Levenez, Alric, & Brugère, 2010; Reeves-Daniel et al., 2010; Woodmansey, McMurdo, Macfarlane, & Macfarlane, 2004). The presence of *M. smithii* has been linked to *Bacteroides* dominated gut microbiomes, while sulfate-reducing *Desulfovibrio* has been linked to *Ruminococcus* dominated gut microbiomes, suggesting further interactions with other microbial communities (Arumugam et al., 2011).

The role of eukaryotic fungi in the gut microbiome has not been thoroughly interrogated. Fungal sequences are underrepresented in annotated reference databases and therefore fungi may be relatively under-detected in sequence-based studies (Underhill & Iliev, 2014). Fungi constitute an estimated 0.1% of the gut microbiota (Issa, Badran, Akl, &

Shehabi, 2011; Qi, Hu, & Zhou, 2005) and are substantially larger than bacteria (~5 μm in diameter; bacteria are ~1 μm in diameter). Like bacteria and archaea, fungal colonization typically occurs at or around birth (Issa et al., 2011; Qi et al., 2005) and fungi can be transmitted through close physical contact (Bougnoux et al., 2006; Pierson, Mehta, Magee, & Mishra, 1995). Fungi may contribute unique metabolic features to gut microbiota/host interactions, and fluctuations in fungal proliferation in response to dysbiosis (abnormal gut microbiota) or diet can have considerable impacts on the immune system (Bull-Otterson et al., 2013, p.; Devkota et al., 2012; Sonnenberg et al., 2012). While most fungi are commensal, they have the potential to cause life-threatening infection, particularly in critically ill or immunocompromised patients (Eggimann, Garbino, & Pittet, 2003).

Recent studies using high-throughput sequencing indicate that more than 50 genera of fungi reside in human guts, with *Saccharomyces*, *Cladosporium,* and *Candida* species being the most abundant (Hoffmann et al., 2013; Iliev et al., 2012). *Candida* is the most well-studied yeast species, and has been cultured from the GI tracts of healthy patients and from sufferers of inflammatory bowel disease (IBD), indicating its presence in the gut during normal and disease states (Odds et al., 2006; Standaert-Vitse et al., 2006). Changes in fungal populations may be due to their reduced abundance relative to bacteria, sensitivity to unknown environmental factors, or alteration in diet. In humans, consumption of an animal based diet has been linked to an increase in *Penicillium* species, whereas consumption of a plant based diet has been linked to an increase in *Candida* species (David et al., 2014).

The human gut virome is constituted primarily by bacteriophages (phages) and, to a lesser extent, by plant/animal/amoebae-infecting viruses. Phages can follow either a lysogenic (latent) or lytic (virulent) lifecycle, and lysogenic viruses can transition to lytic viruses in the presence of an environmental stressor (e.g., antibiotics). Phage populations are generally specific to individuals (Castro-Mejía et al., 2015, p.; Hoyles et al., 2014), and relatively stable over time (Broecker, Russo, Klumpp, & Moelling, 2017), although they can rapidly accumulate sequence variation (Minot et al., 2013). Identifiable prophages in bacterial genomes indicate that their predicted hosts are members of the *Firmicutes* and *Bacteroidetes* phyla, which are the most abundant bacteria in the human GI tract (Carding, Davis, & Hoyles, 2017). Metagenomic sequencing has become integral to human virome characterization given the high genetic diversity of viruses and lack of a universal marker gene. However, only a small fraction (~4-22%) of gut microbiome reads will map to viral DNA unless extraction steps are taken to enrich on virus-like particle (VLP) targets (Dutilh et al., 2014; Minot et al., 2013; Ogilvie et al., 2013). In general, the fecal viromes of healthy adults tend to be composed of phage in the order of *Caudovirales* (10^{15} phages in total) (Castro-Mejía et al., 2015; Dalmasso, Hill, & Ross, 2014; Hoyles et al., 2014; Lepage et al., 2008), eukaryotic ssDNA viruses (e.g., anelloviruses), eukaryotic dsDNA viruses (e.g., adenoviruses, polyomaviruses), human RNA viruses (e.g., retroviruses), and plant RNA viruses (e.g., *Pepper mild mottle virus),* the latter of which are transient members of the human gut virome (e.g., ingested from food) (Carding et al., 2017).

Protists, including protozoan parasites, helminths, and amoebas, can be commensal or opportunistic pathogens of the human gut (Laforest-Lapointe & Arrieta, 2018). The degree to which protists within the gut could affect or be affected by surrounding microbiota is not fully understood. Protists in the gut or in other tissues may also affect the gut microbiota indirectly via the immune system and host metabolism. A detailed discussion on putative protist-microbiome interactions is provided in the sub-section of 1.3 entitled, "The role of parasitemia and HIV in the gut microbiome."

The advent of next-generation metagenomic sequencing

Many microbes in the gut microbiome are difficult to culture (Amann, Ludwig, & Schleifer, 1995), and thus early culture-based studies of gut microbiome composition underestimated its diversity. Sanger sequencing was used to characterize microbial DNA in early microbiome studies (Collado, Donat, Ribes-Koninckx, Calabuig, & Sanz, 2009), but has been largely supplanted by next-generation sequencing (NGS) (Collado et al., 2009). NGS allows targeted amplification of hundreds of thousands of microbial samples using multiplexed "barcodes" to tag fragmented sample DNA (Hamady, Walker, Harris, Gold, & Knight, 2008). The top 99.9% of bacteria can be characterized using this method, and no prior knowledge of the sequence data is required besides the 16S ribosomal RNA (16S rRNA) gene primer sequences. Furthermore, NGS can quantitatively measure mRNA transcripts (or the "transcriptome") to assess microbial gene expression.

The most commonly employed NGS strategies are targeted amplicon sequencing and untargeted shotgun sequencing. Conserved marker genes, functional genes, non-coding

sequences, and 16S rRNA genes can be used in phylogenetic classification, although 16S rRNA is currently the most commonly analyzed gene. The 16S rRNA gene is a 1.5kb gene in the 30S small subunit of prokaryotic ribosomes. It is conserved between different species of *Archaea* and *Bacteria* (Coenye & Vandamme, 2003) and has nine hypervariable regions that contain species-specific sequences that allow bacterial identification. Each of these regions is phylogenetically informative and can be targeted with sequence-specific primers. The most commonly used regions are V1-V6 due to their longer fragment length (informative for identifying genera) and overlapping reads (V1/V2 and V3/V4 paired region sequencing reduces noise in data and genera inflation)(Marchesi, 2014). To classify archaea, the V6-V9 sections of the 16S rRNA gene are frequently used (Hoffmann et al., 2013; McKenna et al., 2008). The 18S rRNA gene or the V6-V9 sections of the internal transcribed spacer (ITS) region in the nuclear ribosomal repeat unit are regularly used to type fungi (Ghannoum et al., 2010; Hoffmann et al., 2013). Enrichment of the target sequences is performed prior to metagenomic sequencing using polymerase chain reaction (PCR) amplification, various enzymatic methods, and hybrid capture approaches. While financially tractable, amplicon sequencing is limited in its ability to describe more than the microbial kingdom of choice, and primer bias for 16S rRNA regions can result in a loss of sensitivity for different taxa of interest (Clooney et al., 2016; Meisel et al., 2016).

Nucleotide sequence data require assembly, taxonomic assignment, and functional annotation of microbial genomes. To avoid misassembled contigs (chimeras) and uneven coverage of low abundance taxa, reference-based assembly and *de novo* assembly are

used. Reference based assembly relies on a mapping the consensus contigs to a reference database of genomes[1] and is limited by the availability of relevant reference genomes in the database. *De novo* assembly requires more computational memory and time, but can produce novel sequences by clustering sequences based on similarity or assigning taxonomies to clusters based on matching representative sequences in a reference database. Taxonomic classification, or "binning," is done in *de novo* assembly by using either: 1) a similarity based method, which places sequences in the same taxonomy based on close relationship to a reference genome, 2) a composition based method, which places sequences in the same taxonomy based on similarity in sequence composition, or 3) an abundance based method, which uses abundance differences between species to classify taxa (Morgan & Huttenhower, 2012).

Functional annotation refers to identifying regions within DNA sequences that encode RNA and/or protein-coding genes. Determining if a sequence is functional (coding vs. noncoding) is very different from determining the function of a gene. Functional annotation can be performed through various open-source software packages such as Quantitative Insights Into Microbial Ecology (QIIME, QIIME2) (Caporaso et al., 2010, https://qiime2.org), Ultra-Fast Sequence Analysis (USEARCH) (Edgar, 2010), and mothur (Schloss et al., 2009).

[1] These include NCBI Microbial Genomes (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi?view=1), the Ribosomal Database Project (http://rdp.cme.msu.edu/), Greengenes (http://greengenes.lbl.gov/), the Human Oral Microbiome Database (http://www.homd.org/) and the HMP Data Analysis and Coordination Center (http://hmpdacc.org/).

Whole-genome shotgun sequencing methods aim to identify all DNA in a sample, independent of its origin, and without a marker-gene amplification step. Sequenced reads are then mapped back to microbial databases. Shotgun metagenomic data allow the reconstruction of partial-to-full microbial genomes, which provides information about the repertoire of genes as well as genomic structure and function (Morgan & Huttenhower, 2012). Highly complex shotgun data can be computationally expensive and time consuming to process as well as difficult to analyze, as reads often cannot be mapped back to reference genomes due to a lack of available information in databases and incomplete genomic characterization. For this same reason, viruses tend to be underrepresented in public databases (Bzhalava, Hultin, & Dillner, 2018).

In general, human metagenomics pipelines involve quality control processing and host read decontamination, followed by multiple downstream steps to classify, assemble, and search reads against existing databases (Peabody, Van Rossum, Lo, & Brinkman, 2015). Currently, many 16S rRNA studies are complemented with whole-genome shotgun sequencing on only a subset of samples (Clemente et al., 2015; Consortium, 2012; Obregon-Tito et al., 2015; Smits et al., 2017; Yatsunenko et al., 2012), often due to cost limitations. Presumably, as sequencing costs decrease, amplicon target enrichment and sequencing will be replaced by whole-genome sequencing strategies.

Origins, establishment, and aging of the human gut microbiome

Comparisons of gut microbiomes of bonobos, chimpanzees, gorillas, and humans revealed that the mean level of microbial diversity in multiple human populations' gut

microbiomes are less diverse than those of non-human primates, supporting the hypothesis that gut microbial diversity has decreased during human evolution (Moeller et al., 2014). An alternative, less parsimonious hypothesis posited by Moeller and colleagues (2014), was that *Pan* and *Gorilla* species have experienced increased gut microbiome diversity since diverging from humans >7 mya (Moeller et al., 2014; Stone et al., 2010). African ape species share bacterial assemblages dominated by *Bacteroides, Prevotella,* and *Ruminococcus* (Arumugam et al., 2011), which may indicate that these microbial relationships in host guts may predate the diversification of African ape species and humans. Phylogenetic reconstructions of African ape and human microbiomes indicate consistent differences arising in host species since their divergence (Moeller et al., 2014). From the Moller et al. study (2014), 35 instances of relative microbial abundance taxon shifts have occurred since the divergence of gorillas and humans, including 17 specific to humans. Of these 17, several may have functional significance for host nutrition, including a more than fivefold increase in relative abundances of *Bacteroides* in humans compared to other apes, which is implicated in digestion of animal fats, and a substantial reduction in *Fibrobacter*, a common plant-fermenting genus characteristic of wild ape microbiomes (Moeller et al., 2014).

Longitudinal ecological studies show that gut microbiota are stable, and that early colonizing microbiota of the gut (including taxa acquired by offspring from parents) have a pivotal role in determining microbiome composition (Faith et al., 2013; Lozupone, Stombaugh, Gordon, Jansson, & Knight, 2012; Yatsunenko et al., 2012). The gut microbiota of pregnant women have been shown to differ between mothers, become

reduced in richness (See Table 2-1), and increase in *Proteobacteria* and *Actinobacteria* taxa between the first and third trimesters (Koren et al., 2012). Furthermore, gut microbiota composition in the third trimester are correlated with weight gain and insulin desensitization. Transfer of third trimester fecal microbiota from mothers into germ free mice induced greater inflammation and adiposity than did first trimester microbiota. Taken together, these lines of evidence show that pregnancy has a profound impact on the gut microbiome of pregnant women and is strongly correlated with changes in host phenotype (Koren et al., 2012).

Metric	**Definition**
Alpha Diversity	The number (richness) and distribution (evenness) of taxa within a single population, i.e., the within-person diversity. Examples include Shannon Index, rarefaction curves, Chao1 index, and Faith's phylogenetic diversity.
Beta Diversity	The absolute or relative amounts of taxa shared between two units of analysis. This acts as a similarity score between populations, i.e., comparing between two populations or individuals. Weighted beta diversity measures population abundance, and unweighted beta diversity measures population presence/absence.

Table 2-1. Commonly used statistical terms in microbiome research and their definitions

Mounting evidence indicates that the placenta is a sterile environment (Lauder et al., 2016; Leiby et al., 2018; Perez-Muñoz, Arrieta, Ramer-Tait, & Walter, 2017), save for some pregnancies with adverse events (e.g., preterm birth) (Ardissone et al., 2014; Payne & Bayatibojakhi, 2014; Prince, Antony, Chu, & Aagaard, 2014). Infant gut microbiomes are influenced by mode of delivery, with some taxa probably transferred vertically from

mother to child. Caesarean-section (C-section) deliveries transmit environmental and skin microbes to the infant, whereas vaginal deliveries transmit vaginal and perianal microbiota (Brumbaugh et al., 2016; Dominguez-Bello et al., 2010; Gilbert, 2014; Song, Dominguez-Bello, & Knight, 2013). Artificial "seeding" of the microbiome for C-section infants using vaginal swabs was suggested to establish a persistent vaginal-like state of early infant gut microbiomes, indicating that critical windows may exist for manipulation of early gut microbiome communities (Dominguez-Bello et al., 2010). Although differences in species richness have been noted for C-section infants compared to those delivered vaginally (Mueller, Bakacs, Combellick, Grigoryan, & Dominguez-Bello, 2015), there remains scant evidence to link C-section associated microbiomes to deleterious health outcomes for infants (Stinson, Payne, & Keelan, 2018). Infant microbiomes also undergo a series of changes with the introduction of solid food regardless of delivery mode that ultimately produce an adult-like microbial composition (Bäckhed et al., 2015; Koenig et al., 2011).

After initial seeding at birth, the main determinant of infant gut microbial composition is feeding mode. Human milk is not sterile, and breastmilk can transmit maternal antibodies, milk microbiota, and skin-associated taxa to infant guts. Breastmilk composition can be affected by multiple factors, including maternal age, diet, metabolic and hormone status, and delivery (Nuriel-Ohayon, Neuman, & Koren, 2016). Exclusively formula-fed babies are enriched in *Streptococcus* and *Enterococcus*, whereas breastfed babies tend to have several skin-associated genera and distinct phylotypes of *Actinobacteria* and *Bifidobacteria* (Timmerman et al., 2017). Human milk

oligosaccharides (HMOs) are one of the most abundant components of breastmilk, and although they are indigestible by host intestinal enzymes, they can be directly hydrolyzed by some species of beneficial colonic *Bifidobacteria* (Marcobal et al., 2010), providing evidence of host-microbial coevolution.

Globally, gut microbiome composition between young adults, elderly, and centenarians does not change linearly with age. Age-associated changes in these groups vary based on population and geographic location (Benno et al., 1989; Hayashi, Sakamoto, Kitahara, & Benno, 2003; Mäkivuokko, Tiihonen, Tynkkynen, Paulin, & Rautonen, 2010; Mueller et al., 2006; Zwielehner et al., 2009). Centenarian microbiomes show high species diversity, a decrease in certain *Clostridia* species and shifts in abundance of several butyrate producing species (Maslowski et al., 2009), which have been linked to longevity. Conversely, pathobiont species of *Proteobacteria* and decreased diversity have been associated with increased inflammation and fragility in elderly groups (Jackson et al., 2016).

Influence of non-dietary factors on the gut microbiome

The composition of an individual's gut microbiome is influenced by several non-dietary factors, including sex, genetics, nutrition status, health, socioeconomic group, geographical location, age, lifestyle, and medication (Deschasaux et al., 2018). Diet is discussed in the next section, "Influence of dietary factors on the gut microbiome."

Conflicting results regarding the effect of sex on the human gut microbiome have been found, with some studies noting differences in taxa by sex, including higher *Bacteroides*

(Mueller et al., 2006), *Clostridia, Proteobacteria* (Li et al., 2008), *Bacteroidetes* (Dominianni et al., 2015; Li et al., 2008), *Veillonella,* and *Methanobrevibacter* (Haro et al., 2016) in men and higher *Bilophila* (Haro et al., 2016) in women. Other studies have found no difference between male and female gut microbiomes (Human Microbiome Consortium, 2012; Dicksved et al., 2007; Lay et al., 2005). Enrichment of taxa based on sexual division of labor and consequent differential nutrient consumption has been noted for certain traditional populations, including an enrichment of *Treponema* in Hadza women and *Blautia* and *Eubacterium* in Hadza men (Schnorr et al., 2014). Hansen et al. (2019) also found sex-specific gut microbiome differences in the Hadza, and additionally found sex differences in the gut microbiomes of pastoralist Maasai (see Chapter 2 for details). Confounding effects including BMI, diet, sex hormones, age, and genetic background, making the discernment of human sex differences in the gut microbiome an ongoing challenge (Bolnick et al., 2014; Elderman et al., 2017; Org et al., 2016; Yurkovetskiy et al., 2013).

Several studies have associated hundreds of host genome-wide loci to microbiome composition, diversity, and taxon abundance (Blekhman et al., 2015; Bonder, Kurilshikov, et al., 2016; Goodrich et al., 2016; Rothschild et al., 2018; Turpin et al., 2016; Wang et al., 2016; Xie et al., 2016), however, the only overlap amongst different studies has been a significant positive association between genetic variants in the *LCT* locus with *Bifidobacteria* abundance (Goodrich, Davenport, Clark, & Ley, 2017). The human *LCT* gene encodes the lactase enzyme that facilitates lactose digestion, and *Bifidobacteria* catabolize milk sugars (i.e, lactose) (Bouhnik et al. 2004). Rothschild et al.

(2018) estimated the average heritability of the gut microbiome of >2k mono- and dizygotic twins to be between 1.9%-8.1%. Unfortunately, host single-nucleotide polymorphism (SNP)-bacteria associations have been population dependent and tend to be underpowered due to small sample sizes and technical artifacts.

Acute and chronic infections with other microbial pathogens can disturb the host gut microbiome, and individual bacterial infections in the GI tract have been identified as the etiological agents behind multiple diseases. *Vibrio cholerae, Yersinia pestis, Mycobacterium tuberculosis* and *Treponema pallidum* have co-evolutionary relationships with humans dating to antiquity, and produce the diseases cholera, plague, tuberculosis, and syphilis, respectively (Spyrou, Bos, Herbig, & Krause, 2019). Introduction of *Helicobacter pylori* was directly attributed to the formation of stomach ulcerations after its infamous ingestion by physicians (Marshall, Armstrong, McGechie, & Glancy, 1985; Morris & Nicholson, 1987). Other bacteria, such as *Escherichia coli* and *Clostridium difficile,* are commensal members of the gut flora that can become toxigenic when the gut microbiome is disturbed from its normal state, i.e., by antimicrobial-induced shifts in bacterial diversity (Goldin & Gorbach, 1980). Declines in bacterial diversity have been detected as much as three years from the initial antibiotic exposure (Leong, Derraik, Hofman, & Cutfield, 2018).

Some pathogenic bacterial infections are a major global public health concern; in 2011, *C. difficile* infections alone produced ~29,000 deaths in the United States due to recurrent, acute diarrhea and colonic inflammation (“CDC Press Releases,” 2016). As various bacteria have become resistant to conventional chemotherapy, fecal

transplantation has been used from healthy donors to restore gut microbiome function and composition in human patients suffering from *C. difficile* associated diarrhea (Aas, Gessert, & Bakken, 2003; Bakken et al., 2011; Khoruts, Dicksved, Jansson, & Sadowsky, 2010; Peterfreund et al., 2012; van Nood et al., 2013).

Although individual bacteria can cause disease, studies of the gut microbiome focus on the contribution of a complex consortium of gut microbes with regard to the predisposition for, acquisition of, and response to disease. Gut microbiota are critical for establishment and regulation of the immune system (Amann et al., 1995; Blekhman et al., 2015; Koren et al., 2012; Million et al., 2012; Renz et al., 2017; Schwiertz et al., 2010).. The increased prevalence of atopic disease in Western societies as a byproduct of increased sanitization, decreased early life microbial exposure, and hypersensitivity of the immune system has been referred to as the "Hygiene Hypothesis" (Isolauri, 2004; Okada, Kuhn, Feillet, & Bach, 2010; Strachan, 2000). This has been augmented by the "Old Friends" hypothesis (Grammatikos, 2008; Rook, Martinelli, & Brunet, 2003), which proposes that ambient microbes co-exist with humans and that these microbes (e.g., helminths, viruses) reside in a carrier or chronic state tolerated by the immune system, with variable amounts of interdependency. Together, these hypotheses have led to speculation that diet and antibiotic use are inhibiting or changing normal development of the immune system (Noverr & Huffnagle, 2005). For example, initial allergic responses often arise in the gastrointestinal tract as a result of food allergens. Once certain foods are ingested, irregular functioning in the gut mucosa can result in greater antigen transfer from the mucosal barrier to the innate immune system, which evokes an aberrant immune

reaction and the production of proinflammatory T helper cell 2-type (Th-2) cytokines (Marchesi, 2014). Th-2 cytokines include interleukin (IL) 4, IL-5, and IL-3, which promote IgE production. Interestingly, Th-2 immune responses may be inversely related to cytokines produced by Th-1, Th-3, and Treg cells, partially as a result of stimulation by the gut microbiota (sometimes called "Th1/Th2 balance") (Rautava, Collado, Salminen, & Isolauri, 2012). Increased proinflammatory cytokines can weaken barrier function of the gut mucosa, potentiating an auto-catalytic cycle of increasing allergic sensitivity and dysregulation in immune response (Marchesi, 2014).

Many studies have described the differences in gut microbiota of infants who developed allergic disorders (increased *Bifidobacterium adolescentis* and *Bifidobacterium pseudocatenulatum)* and healthy infants without allergic disorders (increased *Bifidobacterium breve, Bifidobacterium longum* subsp. *infantis, Bifidobacterium bifidum*) (Björkstén, Sepp, Julge, Voor, & Mikelsaar, 2001; Kalliomäki et al., 2001; Penders et al., 2007, 2006; Ventura, Canchaya, Fitzgerald, Gupta, & van Sinderen, 2007). The gut microbiota of infants with allergic disorders more closely reflects the bacterial types found in adult gut microbiomes (Ventura et al., 2007). Table 2-2 summarizes associations between immune and metabolic features with changes in particular bacterial taxa.

Immune, Metabolic Component	Pre-biotics	Pro-biotics	Poly-phenols	Unsaturated fat	Saturated fat	Animal protein	Pea protein	Artificial sweeteners
C-reactive Protein (CRP)		↓	↓					
High density lipoprotein (HDL)		↑	↑					
Immunoglobulin A (IgA)		↑	↑					
Insulin sensitivity	↑	↑			↓			↓
Insulin-like growth factor -1 (IGF-1) production						↑		
Interleukin-6 (IL-6)	↓							
Interleukin 10 (IL-10)	↑	↑						
Lipopolysaccharides (LPS)	↓			↓	↑			
Low density lipoprotein (LDL)	↓	↓		↓				
Metabolic Endotoxemia				↓	↑			
Plasma Triglycerides		↑	↑					
Short-chain fatty acids (SCFAs)	↑	↑				↓	↑	
Toll-like receptor activation (TLR)				↓	↑			
Total cholesterol	↓	↓		↓				
White adipose tissue (WAT)				↓	↑			

Table 2-2. Effects of dietary components on immune and metabolic parameters; Adapted from Singh et al. (2017).
References: **(Akatsu et al., 2013; Caesar, Tremaroli, Kovatcheva-Datchary, Cani, & Bäckhed, 2015; Cani et al., 2008; M. Carmen Collado, Surono, Meriluoto, & Salminen, 2007; Cuervo et al., 2016; Druart et al., 2014; Filippis et al., 2016; Foligné et al., 2016; Francavilla et al., 2012; Inoguchi et al., 2012; Jantchou, Morois, Clavel-Chapelon, Boutron-Ruault, & Carbonnel, 2010; Kang et al., 2010; Keim & Martin, 2014; M.-S. Kim, Hwang, Park, & Bae, 2013; Lecomte et al., 2015; Levine et al., 2014; J.-E. Liu et al., 2010; Machiels et al., 2014; F.-Martin et al., 2012; Martínez et al., 2013; Matsumoto et al., 2010; Queipo-Ortuño et al., 2012; Rajkumar et al., 2014; Schley & Field, 2002; Singh et al., 2017; Suez et al., 2014; Świątecka et al., 2011; Tzounis et al., 2008; Urwin et al., 2014; S. Wang et al., 2012; West et al., 2013; Yu, Liu, Mukherjee, & Newburg, 2013)**

Shared environments between humans and animals such as dogs have been associated with an increase in overlapping bacterial OTUs, although dogs still maintain gut microbiomes distinct from cohabitating humans (Song, Lauber, et al., 2013). Early exposure to dogs and farm animals has been tied to a decrease in the risk of atopic diseases in children, and animals may increase bacterial diversity in children's gut microbiomes (Fall et al., 2015; Kettleson et al., 2015; Torres et al., 2017). This observation indicates that cross-species transmission of microbiota is not only possible, but may modulate bacterial diversity and help attenuate immunity in industrialized, sterilized environments.

Besides atopic illness, autoimmune diseases including Type 1 diabetes, IBD (Crohn's disease, ulcerative colitis), celiac disease, as well as nonalcoholic fatty liver disease (NAFLD), inflammatory state nonalcoholic steatohepatitis (NASH), colon cancer, metabolic disease, and obesity, have all been associated with dysbiosis (Kao, Hotte, Gillevet, & Madsen, 2014; Kau, Ahern, Griffin, Goodman, & Gordon, 2011; Wu et al., 2010). Type 1 diabetes is characterized by insulin deficiency and results from autoimmune destruction of pancreatic beta cells. Increasing evidence points towards the

gut microbiota having a role in the progression of Type 1 diabetes. Studies in rat models of diabetes showed that the microbiota of diabetes-resistant rats were higher in *Bifidobacterium* and *Lactobacillus* (Roesch et al., 2009). The pathogenesis of IBD has been associated with a loss of microbial community diversity and an increase in pathogenic species including *Bacteroides vulgatus*, *E. coli*, and *C. difficile* (Meyer, Ramzan, Loftus, Heigh, & Leighton, 2004). While certain alleles in the CARD15 gene have been associated with intestinal inflammation and an account for up to 20% of the risk for Crohn's disease (Guarner & Malagelada, 2003; Henckaerts, Figueroa, Vermeire, & Sans, 2008), these changes do not fully explain the mechanism of immune dysregulation. The success of fecal transplants from healthy donors to treat Crohn's disease also indicates that the gut microbiome may play a significant role in IBD (Bak et al., 2017; Kao et al., 2014). Preclinical evidence also supports a role for gut microbiota in the establishment and progression of liver diseases like NAFLD and NASH, although the mechanisms mediating the gut-liver axis are still poorly understood (Bashiardes, Shapiro, Rozin, Shibolet, & Elinav, 2016). Celiac disease is an inflammatory auto-immune disease triggered by the proteins of rye, wheat, and barley. Fecal microbial analysis of celiac patients showed an increase in proportions of *Bacteroidetes/Prevotella* and a reduction in *Clostridium histolyticum, Clostridium lituseburense, Faecalibacterium prausnitzii*, and *Bifidobacterium* spp. (Cenit, Olivares, Codoñer-Franch, & Sanz, 2015; De Palma et al., 2010; Nadal et al., 2007).

While there has been little indication that the microbiome directly affects immune response against cancer, it has been implicated in the onset of colorectal carcinoma

(CRC). Rats with commensal bacteria that promoted gastrointestinal inflammation after exposure to carcinogens develop colorectal carcinomas, while germ-free rats exposed to the same agents did not (Uronis & Threadgill, 2009). Lactic acid bacteria have also been observed to prevent carcinogenic tumors and lesions in mice (Goldin & Gorbach, 1980; Goldin, Gualtieri, & Moore, 1996; Rowland, Rumney, Coutts, & Lievense, 1998). Specific bacteria (*Fusobacterium nucleatum, E. coli, Bacteroides fragilis)* have been experimentally shown to promote intestinal tumorigenesis in CRC patients (Tilg, Adolph, Gerner, & Moschen, 2018).

Whether obesity and metabolic disease (a group of risk factors including high blood sugar and blood pressure, abnormal triglyceride or cholesterol levels, and increased abdominal fat) causes or are affected by altered gut microbial communities is widely debated. A few studies have described a relative decrease (50%) in the number of *Bacteroidetes* and an increase in *Firmicutes* in genetically rendered obese mice (Ley et al., 2005) and overweight human subjects (Ley, Turnbaugh, Klein, & Gordon, 2006). *Akkermansia mucinophila* has garnered recent attention for its ability to reduce insulin resistance, fat mass, and dyslipidemia in mice (Plovier et al., 2017) and humans (Dao et al., 2016). Microflora from obese human subjects have shown an enrichment for genes associated with carbohydrate and lipid metabolism (Turnbaugh & Gordon, 2009). Other studies have indicated that the opposite trend is true (Schwiertz et al., 2010), or that obesity most closely correlates with shifts in different bacterial taxa (*Bifidobacteria, Methanobacteria, Lactobacilli)* (Million et al., 2012). Mice fed high fat diets showed higher levels of lipopolysaccharides and *Enterobacteriaceae*, which resembles the

metabolic state associated with obesity and insulin resistance. This finding points to obesity as the cause and not the result of altered gut microbiomes (Cani et al., 2007). Furthermore, genetically engineered mice lacking the gene for Toll-like receptor 5, an important innate immune system receptor, were found to eat 10% more than wild type mice and show signs of metabolic syndrome, including insulin resistance, increased adiposity, hypertension, and substantial changes in *Firmicutes* and *Bacteroidetes* taxa (Vijay-Kumar et al., 2010). These outcomes were mirrored by similar findings from healthy, overweight, and obese humans (Fernandes, Su, Rahat-Rozenbloom, Wolever, & Comelli, 2014).

The connection between immune function, gut microbiome changes, and phenotype may also indicate a link to host genotype. Overall, the primary driver of obesity still appears to be a positive energy balance (increase of calories relative to energy expenditure), and reducing caloric intake while increasing energy expenditure can reverse obese phenotypes and the associated alterations in gut microbial communities (Ley et al., 2005). Nevertheless, the transmission and manipulation of gut microbiota provides an intriguing therapeutic target for the prevention and treatment of chronic conditions and diseases.

Influence of dietary factors on the gut microbiome

The composition of large intestine microbial communities can vary substantially in healthy individuals in as little as 24 hours (David et al., 2014; Turnbaugh & Gordon, 2009; Wu et al., 2011), although the proportions and absolute numbers of different taxa within a single person remain relatively stable over time. This observation has prompted the question of whether a "core microbiome" maintains gut homeostasis. This idea has

been bolstered by studies showing that the human gut microbiome is enriched for genes coding for metabolic pathways involved in energy conservation and biosynthesis of amino acids, nucleotides, vitamins, secondary metabolites, and carbohydrates (Gosalbes et al., 2011; Turnbaugh, Henrissat, & Gordon, 2010; Verberkmoes et al., 2009). In a meta-analysis of nearly 4000 individuals from multiple human populations distributed across the world, Falony et al. (2014) found 14 "core genera"[2] comprising a global human core microbiome. The cumulative effect size of the global core microbiota on total microbial variation across multiple populations was only 7.63%, emphasizing that total gut diversity hasn't been fully characterized (Falony et al., 2016).

Three general gut microbial "enterotypes", or classifications of gut microbiota based on their bacterial communities, have been proposed for humans based on the relative abundance of bacteria present: *Bacteroides* (enterotype 1), *Prevotella* (enterotype 2), and *Ruminococcus* (enterotype 3) (Arumugam et al., 2011). Wu et al. (2011) found that these enterotypes are associated with long-term dietary patterns and bacterial nutrient-processing preferences: *Prevotella* was predominant in individuals on vegetarian diets with carbohydrate enriched metabolism, while *Bacteroides* was associated with omnivorous diets high in animal fats and protein. *Ruminococcus* was associated with fat enriched "Western" diets (De Filippo et al., 2010). However, more recent studies have failed to find associations between enterotype and diet (Wu et al., 2014), or have found

[2] The core genera were *Roseburia, Faecalibacterium, Dorea, Coprococcus, Clostridium XIVa, Blautia, Bacteroides*, and unclassified taxa in the following families: *Veillonellaceae, Ruminococcaceae, Lachnospiraceae, Hyphomicrobiaceae, Erysipelotrichaceae, Clostridiales*, and *Clostridiaceae* (Falony et al., 2016).

alternate, strong associations between enterotype and ethnicity (He et al., 2018), enterotype and geography (Deschasaux et al., 2018), and enterotype with industrialization (Ayeni et al., 2018; Gomez et al., 2016; Stagaman et al., 2018). Thus, the factors shaping enterotype composition are not fully clarified.

Abundance of bacterial taxa can increase or decrease in response to different dietary substrates. Dietary variation can come from differences in local environment, genetic factors in host metabolism, cultural practices surrounding food processing and consumption, and individual preference. For instance, gut microbiomes of Japanese individuals contain a specific strain of *Bacteroides plebeius* that has acquired a novel gene from marine bacteria which degrades the polysaccharide porphyran in seaweed, a common component of many Japanese diets (Hehemann et al., 2010). This gene was not detected in North American gut microbiomes, and demonstrates adaptation to local diet by the gut microbial community. Some bacteria are better able to catabolize particular food-derived sugars and fats over others. A summary of the associations between bacterial abundance and microbiome features (e.g., microbial diversity) for different dietary components is provided in Table 2-3. Several studies have analyzed the effects on the gut microbiome of different diets, including vegan, gluten-free, vegetarian, Mediterranean (high in antioxidants/fiber/unsaturated fatty acids/low in red meats), omnivorous, and "Western" (high in protein/animal fat) (Table 2-3). By contrast, recent studies of industrialized U.S. and Italian cohorts indicated that omnivores and vegans showed negligible gut microbiome differences (Wu et al., 2016) which may be attributed to their equal intake of nutrients, regardless of food source (Losasso et al., 2018).

Taxa or Feature	High Fat	Low Fat	High Saturated Fat	High Unsaturated Fat	Animal Protein	Whey Protein Extract	Pea Protein Extract	Glucose
Akker-mansia muciniphila				↑				
Alistipes					↑			
Bacteroides	↑		↑		↑↓	↓		↓
Bifido-bacteria		↑		↑	↑↓	↑	↑	↑
Bilophila			↑		↑			
Clostridia					↑	↓		
Clostridiales	↑							
Entero-coccus					↑↓			
Eubacterium rectale					↑↓			
Faecali-bacterium prausnitzii			↑					
Lactic acid Bacteria	↓			↑				
Lactobacilli						↑	↑	
Microbial Diversity					↑	↑	↑	
Roseburia					↓			

Taxa or Feature	Fructose	Sucrose	Lactose	Artificial Sweeteners	Fiber/ prebiotics	Resistant Starch
Bacterial Abundance					↑	↑
Bacteroides	↓	↓	↓	↑		

Bifidobacteria	↑	↑	↑	↓	↑	↑
Bilophila			↓	↓		
Clostridia					↓	
Enterococcus					↑↓	
Eubacteria						↑
Gene Enrichment					↑	↑
Helicobacter pylori						
Lactobacilli			↑	↓	↑	↑
Roseburia						↑
Ruminococcus						↑
Streptococcus						↑
Total aerobes/ anaerobes						↑

Taxa or Feature	**Probiotics**	**Polyphenols**	**"Western" diet**	**Mediterranean diet**	**Gluten-Free Diet**
Bacterial Abundance	↑				
Bacteroides		↓	↑	↑	
Bifidobacteria	↑	↑	↓	↑	↓
Clostridia		↓			
Enterobacteria			↑		↑
Escherichia coli	↓				
Eubacteria			↓	↑	↓
Helicobacter pylori	↓				
Lactobacilli	↑	↑	↓	↑	↓

Prevotella			↓	↑	↓
Roseburia				↑	↓
Streptococcus	↑				
Total bacteria				↓	
Total coliforms	↓				

Table 2-3. Effects of fats, proteins, carbohydrates, nondigestible carbohydrates, probiotics, polyphenols, and specific diet on gut microbiota. Modified from Singh et al. (2017)
Columns highlighted in orange are fats, columns in pink are proteins, columns in green are carbohydrates, columns in blue are non-digestible carbohydrates, columns in yellow are specific diets. *References*: (Barroso et al., 2014; Bartram et al., 1994; Bialonska et al., 2010, 2010; Bonder, Tigchelaar, et al., 2016; Bouhnik et al., 1996; Cani et al., 2008; Carvalho-Wells et al., 2010, p.; Costabile et al., 2012, 2008; Cotillard et al., 2013; Cuervo et al., 2016, 2014; Cueva et al., 2013; David et al., 2014; De Filippo et al., 2010; De Palma, Nadal, Collado, & Sanz, 2009; De Palma et al., 2009; Del Chierico, Vernocchi, Dallapiccola, & Putignani, 2014; Drasar et al., 1973; Druart et al., 2014; Eeckhaut et al., 2013; Eid et al., 2014; Fava et al., 2013; Filippis et al., 2016; Flickinger et al., 2002; Francavilla et al., 2012; François et al., 2014; García-Albiach et al., 2008; Gomez et al., 2016; Goossens, Jonkers, Russel, Stobberingh, & Stockbrügger, 2006; Gori et al., 2011; Halmos et al., 2015; He et al., 2008; Inoguchi et al., 2012; Jantchou et al., 2010; J.-S. Jin, Touyama, Hisada, & Benno, 2012; Kapiki et al., 2007; Kedia, Vázquez, Charalampopoulos, & Pandiella, 2009; Keim & Martin, 2014; Kim, Park, & Kim, 2014; Kris-Etherton, Harris, Appel, & American Heart Association. Nutrition Committee, 2002; Lecomte et al., 2015; Lee, Jenner, Low, & Lee, 2006; Leitch, Walker, Duncan, Holtrop, & Flint, 2007; Link-Amster, Rochat, Saudan, Mignot, & Aeschlimann, 1994; Liu et al., 2014; Lopez-Legarrea, Fuller, Zulet, Martinez, & Caterson, 2014; Lorenzo Pisarello, Vintiñi, González, Pagani, & Medina, 2015; Machiels et al., 2014; Matsumoto et al., 2010; Meddah et al., 2001; Parvin et al., 2015; Queipo-Ortuño et al., 2012; Rajkumar et al., 2014; Reddy, Weisburger, & Wynder, 1975; Romond et al., 1998; Sairanen et al., 2007; Sánchez-Patán et al., 2012; Spanhaak, Havenaar, & Schaafsma, 1998; Suez et al., 2014; Świątecka et al., 2011; Tormo Carnicer, Infante Piña, Roselló Mayans, & Bartolomé Comas, 2006; Tzounis et al., 2011, 2008; Urwin et al., 2014; Vendrame et al., 2011; Wacklin et al., 2014; Walker et al., 2011; S. Wang et al., 2012; Wu et al., 2011; Yang & Sheu, 2012; Yu et al., 2013; Zhong, Huang, He, & Harmsen, 2006)

Distinguishing taxa that represent resident gut microbes from transient microbes linked to diet is an ongoing area of research. For instance, although fungal DNA in food appears to be mostly degraded during digestion, trace amounts of *Agaricus* sequences from white button mushrooms have been found in human gut microbiomes (Hoffmann et al., 2013).

This observation indicates that some fungal sequences found in gut microbiomes may be acquired from transient foods instead of coming from long term "residents." Hoffman and colleagues (2013) noted that the high prevalence of *Saccharomyces* in their fungal sequences could be due to the ingestion of foods containing yeast, such as beer and bread (Hoffmann et al., 2013).

Microbiome targeted therapies including the administration of prebiotics (dietary supplements that stimulate survival and expansion of probiotics), probiotics (cultured microbiota with putative health benefits), and synbiotics (combinations of pro- and prebiotics) can be considered forms of dietary supplementation or replacement. Various probiotics purported to have beneficial effects within the gut microbiome have been developed from sources including yogurt starter cultures (*B. longum),* infant stool (*B. infantis),* and a World War I soldier who proved resistant to dysentery (*Escherichia coli* Nissle) (Spiller, 2008). A few clinically validated studies have reported improvement for Irritable Bowel Syndrome (IBS), a GI disease associated with disruption of the gut microbiota (Sonnenborn & Schulze, 2009), with use of *B. infantis* probiotics (Brenner, Moeller, Chey, & Schoenfeld, 2009). Other bacteria (*Lactobacillus paracasei*, *Lactobacillus rhamnosus*, and *Bifidobacterium animalis*) have been linked to weight loss and improved glucose-insulin homeostasis (Wang et al., 2015). Particular prebiotics (e.g., oligofructose, inulin) may alter the structural composition of the microbiota that results from a high-fat diet, which could improve inflammation associated with metabolic syndrome and obesity (Nicolucci et al., 2017).

Both the richness and composition of human gut microbiota are altered with consumption of a high-fat diet (Martinez, Leone, & Chang, 2017). Supplementation with dietary fiber improved metabolic syndrome in animals and humans, potentially by increasing microbial short chain fatty acid production (Rosenbaum, Knight, & Leibel, 2015; Sonnenburg & Bäckhed, 2016). The loss of dietary fiber in Western population diets has been linked to a decrease in overall microbial diversity and a rise in chronic conditions, including metabolic disease (Sonnenburg & Bäckhed, 2016). Therapeutic manipulations of the gut microbiota to treat and manipulate obesity and metabolic syndrome include fecal microbiome transplant and antibiotic usage (de Groot, Frissen, de Clercq, & Nieuwdorp, 2017). Metabolic syndrome patients receiving transfers of lean donor stool showed a decrease in fasting triglyceride levels and increased insulin sensitivity (Vrieze et al., 2012). Small observational studies indicate that decreased microbial diversity resultant from human antibiotic treatment is correlated with risk for increased weight gain (Jakobsson et al., 2010; Jernberg, Löfmark, Edlund, & Jansson, 2007, 2010; Panda et al., 2014; Vrieze et al., 2014; Zaura et al., 2015).

2.3 The Gut Microbiomes of Non-industrialized Populations

Expanding the spectrum of normal human gut microbiome variation from study of non-industrialized populations

Changing subsistence practices during the Neolithic demographic transition (~12 kya) are significant factors in recent human evolution (Richerson, Boyd, & Henrich, 2010). The shift from hunter-gatherer diets characterized by starch-rich bulbs and roots to the high starch plant foods and dairy products associated with agriculture and animal

domestication has been suggested to be a major influence in shaping the gut microbiome of AMH (Walter & Ley, 2011). Dairying practices in Neolithic pastoralist groups are strongly correlated with a selective genetic sweep across European and African populations at loci that impact the lactase persistence (LP) trait (Bersaglieri et al., 2004; Tishkoff, Reed, et al., 2007), and host genomic variation at the *LCT* locus that confers the LP trait has been associated with lactose-digesting *Bifidobacteria* in the gut (Goodrich et al., 2017). Many traditional populations practice small-scale agriculture, and cultivate regionally specific crops in community or family-owned gardens, while other populations practice multiple subsistence practices, which can allow access to additional nutritive resources during periods of food scarcity. Historically, the majority of human gut microbiome research has centered around populations from western, educated, industrialized, rich, and democratic countries (Gupta, Paul, & Dutta, 2017; Morton et al., 2015) that practice commercially-sustained food production ("industrial agropastoralism"). It is only recently that the gut microbiomes of diverse, non-industrialized populations have been intensively researched (Ayeni et al., 2018; De Filippo et al., 2010; Gomez et al., 2016; Morton et al., 2015; Rampelli et al., 2015; Schnorr et al., 2014; Smits et al., 2017; Yatsunenko et al., 2012).

Studies on the gut microbiomes of traditional populations have produced a wealth of insights into defining the spectrum of normal and diseased human microbial states (Table 2-4). Metagenomic analyses from Tanzanian Hadza hunter-gatherer gut microbiomes showed an enrichment of carbohydrate metabolizing microbes, which could be an adaptation to their diet which is high in tubers (Rampelli et al., 2015). In CAHG

populations from Cameroon, Morton et al. (2015) found associations between infection with gut parasites and particular gut microbial composition (discussed in more detail in the next section). Longitudinal studies of populations practicing hunting and gathering (Smits et al., 2017) and small scale agropastoralism (Davenport et al., 2014) revealed that recurrent shifts in taxa can be detected across seasons, when different foods are readily available. Despite no known exposure to antibiotics, South American hunter-gatherers possessed an enrichment of functional antibiotic genes, including genes conferring resistance to synthetic antibiotic drugs (Clemente et al., 2015), which could be the indirect result of industrialization through shared environments with overlapping, more settled groups, or due to the presence of antibiotic genes environmentally in high frequencies for reasons unrelated to human antibiotic use..

Population	**Location (Region, Country, City)**	**Subsistence**	**Data Type (Amplicon Sequenced)**	**Study**
Bantu	Africa, Botswana	Rural farming	16S rRNA (V1-V2)	Hansen and Rubel et al. 2019
Herero	Africa, Botswana	Pastoralism	16S rRNA (V1-V2)	Hansen and Rubel et al. 2019
San	Africa, Botswana	Rural farming, hunting and gathering	16S rRNA (V1-V2)	Hansen and Rubel et al. 2019
Burkinabe	Africa, Burkina Faso	Rural farming	16S rRNA (V5-V6)	Filippo et al. 2010
Cameroon hunter-gatherers	Africa, Cameroon	Hunting and gathering	16S rRNA (V5-V6)	Morton et al. 2015
Bantu	Africa, Cameroon	Rural farming	16S rRNA (V5-V6)	Morton et al. 2015
Bantu	Africa, Cameroon	Fishing	16S rRNA (V5-V6)	Morton et al. 2015
BaAka	Africa, Central African Republic	Mixed hunting and gathering, rural farming, trade	16S rRNA (V1-V3)	Gomez et al. 2016
Bantu	Africa, Central African Republic	Rural farming, market agriculture	16S rRNA (V1-V3)	Gomez et al. 2016

Liberians	Africa, Liberia	Rural	Shotgun (subset), 16S rRNA (V1-V3)	Rosa et al. 2018
Malawians	Africa, Malawi, Chamba/Makwhira/Mayaka/ Mbiza	Rural farming, Market Agriculture	Shotgun (subset), 16S rRNA (V4)	Yatsunenko et al. 2012
Bassa	Africa, Nigeria	Rural farming	16S rRNA (V3-V4), Metabolomics (subset)	Ayeni et al. 2018
Southwest Nigerians (Hausa, Igbo, Yoruba, Ebira)	Africa, Nigeria	Urban, market agriculture	16S rRNA (V3-V4), Metabolomics (subset)	Ayeni et al. 2018
Rural South Africans	Africa, South Africa	Rural farming, market agriculture	16S rRNA pyrosequencing, qPCR	Ou et al. 2013
Burunge	Africa, Tanzania	Rural farming	16S rRNA (V1-V2)	Hansen and Rubel et al. 2019
Hadza	Africa, Tanzania	Hunting and gathering	16S rRNA (V4)	Schnorr et al. 2014
Hadza	Africa, Tanzania	Hunting and gathering	Shotgun	Rampelli et al. 2015
Hadza	Africa, Tanzania	Hunting and gathering	Shotgun (subset), 16S rRNA (V4), Metabolomics (subset)	Smits et al. 2017
Hadza	Africa, Tanzania	Hunting and gathering	16S rRNA (V1-V2)	Hansen and Rubel et al. 2019
Maasai	Africa, Tanzania	Pastoralism	16S rRNA (V1-V2)	Hansen and Rubel et al. 2019
Sandawe	Africa, Tanzania	Rural farming, some hunting and gathering	16S rRNA (V1-V2)	Hansen and Rubel et al. 2019
Ugandans (HIV+/-)	Africa, Uganda	Rural farming	16S rRNA (V4), Virome (VLP NGS)	Monaco et al. 2016
Chinese Mongolians	Eurasia, China, Hohhot	Urban, market agriculture	Shotgun	Liu et al. 2016
Han (low altitude)	Eurasia, China, Sichuan Province, Chengdu	Urban, market agriculture	16S rRNA (V1-V3)	Li et al. 2016
Chinese Mongolians	Eurasia, China, Xilingol pasturing area	Pastoralism	Shotgun	Liu et al. 2016
Khentii	Eurasia, Mongolia	Pastoralism	16S rRNA (V1-V3)	Zhang et al. 2014
Khentii	Eurasia, Mongolia	Pastoralism	Shotgun	Liu et al. 2016
Ulan Bator	Eurasia, Mongolia, TUW province	Urban, market agriculture	16S rRNA pyrosequencing, 16S qPCR	Zhang et al. 2014
Ulan Bator	Eurasia, Mongolia, TUW province	Urban, market agriculture	Shotgun	Liu et al. 2016

Rural Russians	Eurasia, Russia, Khakassia/Omsk Tatarstan/Tyva	Rural farming, market agriculture	Shotgun	Tyakht et al. 2013
Urban Russians	Eurasia, Russia, Novosibirsk/Rostov-on-don/Saratov/St. Petersburg	Urban, market agriculture	Shotgun	Tyakht et al. 2013
Han (high altitude)	Eurasia, Tibet, Lhasa	Urban, market agriculture	16S rRNA (V1-V3)	Li et al. 2016
Tibetan herders (high altitude)	Eurasia, Tibet, Qinghai-Tibet plateau	Pastoralism	16S rRNA (V1-V3)	Li et al. 2016
Tibetan herders (low altitude)	Eurasia, Tibet, "Lowlands"	Urban, market agriculture, rural farming	16S rRNA (V1-V3)	Li et al. 2016
Metagenomics of the Human Intestinal Tract (MetaHIT)	Europe, Denmark & Spain	Urban, market agriculture	16S rRNA gene	MetaHIT consortium
TwinsUK	Europe, United Kingdom	Urban, market agriculture	16S rRNA (V4)	Goodrich et al. 2016
TwinsUK	Europe, United Kingdom	Urban, market agriculture	Shotgun	Xie et al. 2016
Urban Italians	Europe, Italy, Bologna	Urban, market agriculture	16S rRNA (V4)	Schnorr et al. 2014
Urban Italians	Europe, Italy, Bologna	Urban, market agriculture	Shotgun	Rampelli et al. 2015
Urban Italians	Europe, Italy, Florence	Urban, market agriculture	16S rRNA (V5-V6)	Filippo et al 2010
LifeLines-DEEP	Europe, Netherlands	Urban, market agriculture	16S rRNA (V4)	Fu et al. 2015
LifeLines-DEEP	Europe, Netherlands	Urban, market agriculture	Shotgun	Zhernakova et al. 2016
Urban Canadians	North America, Canada, Montreal	Urban, market agriculture	16S rRNA (V4)	Girard et al. 2017
Inuit	North America, Canada, Nunavut	Hunting, Fishing, market agriculture	16S rRNA (V4)	Girard et al. 2017
American Gut Project (AGP)	North America, USA	Urban, market agriculture	Shotgun, 16S rRNA (V4), Metabolomics	AGP consortium
Human Microbiome Project (HMP)	North America, USA	Urban, market agriculture	Shotgun, 16S rRNA, Whole-genome sequencing, Transcriptomics, Proteomics, Metabolomics	HMP consortium
Urban USA	North America, USA, Colorado/Boulder, Missouri/St. Louis, Pennsylvania/ Philadelphia	Urban, market agriculture	Shotgun (subset), 16S rRNA (V4)	Yatsunenko et al. 2012
Urban USA	North America, USA, Nebraska, Lincoln	Urban, market agriculture	16S rRNA (V5-V6)	Martínez et al. 2015

Ancient Native American	North America, USA, Northern Mexico, Rio Zape	Hunting and gathering, rural farming	16S rRNA (V3)	Tito et al. 2012
Urban USA	North America, USA, Oklahoma, Norman	Urban, market agriculture	Shotgun, 16S rRNA (V4)	Obregon-Tito et al. 2015
Urban USA	North America, USA, Pennsylvania, Philadelphia	Urban, market agriculture	16S rRNA (V4)	Hansen and Rubel et al. 2019
Hutterites	North America, USA, South Dakota	Rural farming	16S rRNA (V4)	Davenport et al. 2014
Shuar	South America, Ecuador	Hunting, fishing, rural farming, small to moderate market agriculture	16S rRNA (V4)	Stagaman et al. 2018
Matses	South America, Peru	Mixed foraging, rural farming, paracultivation	Shotgun, 16S rRNA (V4)	Obregon-Tito et al. 2015
Amerindian	South America, Venezuela, Bolivar, Kanarakuni Village	Hunting and gathering	16S rRNA (V4)	Ruggles et al. 2018
Yanomami	South America, Venezuela, High Orinoco state	Mixed foraging and paracultivation	Shotgun (subset), 16S rRNA (V4)	Clemente et al. 2015
Guahibo	South America, Venezuela, Platanillal/Coromoto	Rural farming	Shotgun (subset), 16S rRNA (V4)	Yatsunenko et al. 2012
Ballabhgarh (low altitude)	South Asia, India	Rural farming	16S rRNA (V1-V5)	Das et al. 2018
Ballabhgarh	South Asia, India	Urban, market agriculture	16S rRNA (V1-V5)	Das et al. 2018
Hmong	South Asia, Thailand	Rural farming	Shotgun (subset), 16S rRNA (V4)	Vangay et al. 2018
Karen	South Asia, Thailand	Rural farming	Shotgun (subset), 16S rRNA (V4)	Vangay et al. 2018
Leh (high altitude)	South Asia, India	Rural farming	16S rRNA (V1-V5)	Das et al. 2018
Tai-Phage, Tea Tribe, Tai-Aiton, Bodo, Karbi, Gond	South Asia, India, Assam region	Rural farming	16S rRNA (V3-V4)	Dehingia et al. 2015
Koya, Nayak, Kolam	South Asia, India, Telangana region	Rural farming	16S rRNA (V3-V4)	Dehingia et al. 2015
Tangkhul, Kuki, Meitei	South Asia, India, Manipur region	Rural farming	16S rRNA (V3-V4)	Dehingia et al. 2015
Nepalia, Bhutia, Lepcha	South Asia, India, Sikkim region	Rural farming	16S rRNA (V3-V4)	Dehingia et al. 2015
Nicobarese	South Pacific, Andaman and Nicobar Islands	Mixed rural farming and foraging, small to moderate market agriculture	16S rRNA (V3), qPCR	Anwesh et al. 2016
Indonesians	South Pacific, Flores Island	Rural	Shotgun (subset), 16S rRNA (V1-V3)	Wammes et al., 2016, Rosa et al. 2018

Asaro	South Pacific, Papua New Guinea	Rural farming	16S rRNA (V5-V6)	Martínez et al. 2015
Sausi	South Pacific, Papua New Guinea	Rural farming	16S rRNA (V5-V6)	Martínez et al. 2015
Urban, unspecified	Unspecified	Urban, unspecified	16S rRNA (V4)	Ruggles et al. 2018

Table 2-4. List of gut microbiome datasets from traditional subsistence, rural populations and comparative and/or frequently referenced industrialized populations
***Originally published separately in Wu et al. 2011, Minot et al. 2013, and Ni et al 2017**

Lifetime exposure to non-industrialized environments and subsistence economy are strong explanatory variables in gut microbiome composition (Obregon-Tito et al., 2015; Schnorr et al., 2014; Smits et al., 2017). Geographically disparate populations living in environments with comparable levels of antibiotics, hygienic resources, and sterile cleaners who also practice the same subsistence (i.e., industrial agropastoralism, hunting and gathering, etc.) possess compositionally similar gut microbiomes (Schnorr, 2018). Ancient human gut microbiome studies show that ancient humans from Rio Zape, Mexico, have similar taxonomic profiles to contemporary traditional and rural human communities (Tito et al., 2012; Tito et al., 2008), further linking subsistence practice to taxonomic profile, and emphasizing the antiquity of some bacteria within human gut microbiomes. In general, the gut microbiomes of traditional populations are higher in *Prevotellaceae,* and, to a lesser extent, *Ruminococcaceae*, both of which garner energy from carbohydrates and mucins. By contrast, industrialized populations have gut microbiomes that are enriched in *Bacteroidaceae,* which is associated with diets high in animal fats and proteins (Arumugam et al., 2011; De Filippo et al., 2010).

To explore bacterial diversity, Deschasaux et al. (2018) compared 16S rRNA sequences from more than two thousand residents of Amsterdam within six ethnic groups- Dutch,

Moroccans, African Surinamese, South-Asian Surinamese, Turks, and Ghanaians. They found that within-individual (alpha diversity) differences could be detected in the gut microbiome. These results were largely recapitulated in microbiome studies of urban and rural Nigerians (Ayeni et al., 2018). He et al. (2018) found estimates of health made from microbiome data were most accurate when informed by the Chinese participants' local geography. These results indicate that ethnicity and geography are important determinants of gut microbiome composition, independent of lifestyle, demographic, and dietary factors (Deschasaux et al., 2018; He et al., 2018).

Several other studies have highlighted that a transition from non-industrialized to industrialized environments and lifestyle is associated with a decrease in gut microbiome diversity (Clemente et al., 2015; De Filippo et al., 2010; Gomez et al., 2016; Obregon-Tito et al., 2015; Schnorr et al., 2014). Gomez et al. (2016) analyzed metabolites and gut microbiota samples from Central African Republic populations, and showed that while the gut microbiomes of CAHG (BaAka) and Bantu generally looked distinct from those of U.S. populations, the settled, agriculturalist Bantu had gut microbiome profiles that were more similar to those of U.S. guts than those of the BaAka (Gomez et al., 2016). The levels of urbanization and economic development associated with decreased diversity and altered inter-individual taxonomic variation contingent on level of settlement in the Ecuadorian Shuar, a hunting and gathering group transitioning away from traditional lifestyle (Stagaman et al., 2018). Species in the *Treponema* genus are common constituents of the gut microbiomes of traditional, rural populations such as the Venezuelan Yanomami (Clemente et al., 2015) and Tanzanian Hadza (Schnorr, 2018),

but are considered “missing” taxa in the gut microbiomes of industrialized countries. Non-pathogenic species of *Treponema* are thought to facilitate digestion of fibrous foods, and the lack of this genus in industrialized countries may be linked to Western diets that are fiber-depleted, or are incapable of sustaining this bacteria due to other factors such as antibiotic use.

Recent work by Vangay et al. (2018) sampled Hmong and Karen Thai U.S immigrant populations and observed decreased diversity contingent on their duration in the U.S. (Vangay et al., 2018). Hmong and Karen Thai immigrant microbiome profiles converged towards those of European Americans within 6-9 months and coincided with weight gain. The loss of diversity was compounded over generational time and microbial losses such as these are likely permanent (Sonnenburg et al., 2016). Given the diversity of traditional population microbiomes, and that traditional populations are being increasingly exposed to dietary and environmental changes consistent with industrialized countries, some researchers have speculated that certain microbial taxa could be lost from human gut microbiomes in perpetuity. This statement has bolstered calls for international consortia to biobank and preserve microbial diversity in repositories that would serve as long-term, secure “vaults” (Bello, Knight, Gilbert, & Blaser, 2018), although issues of indigenous access, consent, and ownership make such storage and concomitant open-ended research problematic.

Health of traditional African populations

While two-thirds of deaths outside of Africa in 2002 were attributed to noncommunicable diseases, 72% of deaths within Africa are caused by communicable diseases including

malaria, HIV/AIDS, tuberculosis, and neglected tropical diseases (Africa, 2006; Hotez, 2014) (Soil-transmitted helminths, leprosy, onchocerciasis, dracunculiasis, *Schistosoma*, etc.). Concerted efforts to control leprosy, poliomyelitis, dracunculiasis, and onchocerciasis have been comparatively more successful than measures targeting tuberculosis, malaria, and HIV/AIDS despite improvements to disease prevention, detection, and treatment techniques (Africa, 2006).

Malaria is a mosquito-borne infectious illness that causes fever, vomiting, headache, and fatigue, and can lead to seizures, coma, and death. There are an estimated 300-500 million clinical cases of malaria in the world every year, 90% of which occur in Africa. (Africa, 2006). Close to 90,000 children die annually of malaria, and malarial-related anemia causes around 10,000 maternal deaths in Africa (Africa, 2006). Forced settlement and forest clearing activities of Pygmy[3] groups within Central Africa have further propagated mosquito communities (Froment, 2014). It has been hypothesized that increased malaria endemicity from forest clearing may have increased the selection pressure in populations from those regions, resulting in genomic adaptations (Myers et al., 2013; Tishkoff et al., 2001).

---

[3] George Schweinfurth (1873) coined the term "Pygmy" when referring to the rainforest-dwelling hunter gatherers of the Congo Basin based on their short stature (Schweinfurth, 1873). Anthropologists used the term "Pygmy" in the *Journal des Africanistes* (2012) as they lacked a more parsimonious term to refer to people of the Congo Basin, who are no longer confined to rainforests, and have subsistence practices other than hunting and gathering (Robillard & Bahuchet, 2012). Due to its global recognition, this paper also uses the term "Pygmy," although its potential pejorative meaning, embedded in phenotypic classification, is duly noted.

African governments acknowledge the heavy disease burden of malaria, and since 2002 have pledged to treat initial onset of symptoms with expanded clinical care, preventative doses of antimalarial drugs, and insecticide treated nets for at least 60% of "vulnerable groups," defined as children <5 years of age and pregnant women (Organization of African Unity, 2000). However, these simple interventions can be hard to implement due to the scale of need. Nets must be re-treated with insecticide every three years and many countries with high malarial burdens do not have the health infrastructure or financial resources to treat or prevent malaria (Africa, 2006). Furthermore, insecticide resistance, expanded mosquito habitat from climate change, and the proliferation of drug-resistant parasites pose major challenges to eradicating malaria in Africa (Myers et al., 2013; World Health Organization, 2013).

Human immunodeficiency virus (HIV) is a retrovirus that causes acquired immunodeficiency syndrome (AIDS), and is characterized by progressive failure of the immune system and increased rates of opportunistic infections and some cancers. Africa contains 60% of all people living with HIV, making it the region of the world most affected by the HIV/AIDS pandemic (Africa, 2006). HIV is predominantly transmitted among heterosexual people in Africa, and 57% of infected individuals are women (Africa, 2006). Many factors contribute to viral incidence, including migrations, lack of education, poverty, stigma, high levels of sexually transmitted infections (STIs), and social instability (Cohen, 2002). Commercial sex and sexual violence also increase incidence rates, and may disproportionately affect young women (Joint United Nations Programme on HIV/AIDS (UNAIDS), 2004). Increasingly affordable antiretroviral

therapy (ART) lowers HIV levels in blood, postpones opportunistic infections, prevents mother-to-child transmission, and increases quality of life (World Health Organization, 2006). The relative isolation of San populations results in a lower HIV/AIDS rate than populations in urban regions (i.e., an adult prevalence rate of HIV/AIDS in urban Botswanans of 37.3% in 2002, as compared to 21.4% for San). However, forced resettlement into areas with higher HIV/AIDS prevalence threatens to substantially increase the infection rate (Ohenjo et al., 2006).

Tuberculosis (TB), a respiratory infection of pathogenic *Mycobacteria* strains, has an ancient history of causing illness in humans, with a current estimate of its origin at ~3 mya (Gagneux, 2012). Although there has been an affordable and safe cure for TB since the 1950s, millions of people die from the disease every year (*Global Tuberculosis Control*, 2010). In Africa, an estimated 2.4 million new cases and 0.5 million deaths from TB occur annually (Africa, 2006). Africa has nine of twenty-two high TB burden countries (Ethiopia, the Democratic Republic of Congo, Nigeria, Kenya, South Africa, Mozambique, Uganda, Zimbabwe, and the United Republic of Tanzania) and eleven of fifteen countries with the highest incidence of TB (Botswana, Lesotho, Malawi, Kenya, Sierra Leone, Namibia, South Africa, Swaziland, Zambia, Uganda, and Zimbabwe) (Africa, 2006). As pathogen resistance to conventional chemotherapies has increased, new strains such as Multiple Drug Resistant (MDR) and Extremely Drug Resistant (XDR) TB are on the rise, causing increased morbidity and requiring expensive antibiotics for treatment (*Global Tuberculosis Control*, 2010; Sharma et al., 2017).

The incidence of tuberculosis has risen in tandem with the HIV/AIDS epidemic, as people who are immunocompromised due to AIDs easily contract tuberculosis. Additionally, treatment with antiretroviral drugs can lyse granulomatous sequellae of latent TB disease, causing an active and transmissible disease state (Lawn, Butera, & Shinnick, 2002). Close supervision of antibiotic administration with directly observed treatment (DOTS) can produce higher TB cure rates, fewer relapses, and prevents drug resistance,- however, African region countries often lack trained health workers to implement treatment plans (Africa, 2006). The Bacille Calmette- Guérin (BCG) vaccination routinely administered to newborns provides protection against some severe forms of TB, but it is not efficacious against pulmonary TB and does little to relieve the TB burden (Africa, 2006).

Neglected Tropical Diseases (NTDs) continue to be common in Africa, generating substantial morbidity, but do not garner much attention or funding to their treatment and prevention. Among these, schistosomiasis and Soil Transmitted Helminths (STH) are some of the most prevalent infectious agents in developing countries (Hotez, 2014). Schistosomiasis is caused by parasitic worms (common species in Africa include *Schistosoma mansoni* and *Schistosoma haematobium)* which infect the intestines or the urinary tract causing diarrhea, abdominal pain, hematuria, and in cases of prolonged infection, infertility, bladder cancer, kidney failure, and liver failure (Chitsulo, Engels, Montresor, & Savioli, 2000). In schistosome endemic regions, many regional African governments have treated at-risk groups with annual doses of the deworming medication praziquantel, but reinfection is common (Hotez et al., 2006). The most common STH

include intestinal parasitic whipworms (*Trichuris trichiura),* roundworms (*Ascaris lumbricoides),* and hookworms (*Necator americanus* and *Ancylostoma duodenale),* which are contracted through contact via eggs in soil, ingestion of eggs in contaminated food, or through skin penetration by infective larvae (Africa, 2006). Symptoms of infection with STH include abdominal pain, diarrhea, anemia (for hookworms) general malaise, and weakness, which can impair growth, learning ability, and work capacity (Africa, 2006). Campaigns within Africa to control STH include giving targeted drug treatment to children, pregnant women, and high risk adults (Hotez et al., 2006). African populations are also infected by protozoans, including *Entamoeba histolytica*, *Cryptosporidium spp.*, and *Giardia*, which account for ~357 million cases of illness (Torgerson et al., 2015).

Intestinal parasites are endemic to the equatorial rainforests of Africa, which are inhabited by many indigenous African groups (Ohenjo et al., 2006). Diarrheal diseases are the main cause of children's growth defects and mortality in Pygmy groups of Central Africa (Froment, 2014) and most Pygmy groups have comparable if not higher levels of intestinal parasites as their agricultural Bantu neighbors (Ohenjo et al., 2006). Clinics are often located in urban centers, limiting Pygmy access to deworming medications (Ohenjo et al., 2006). The rate of helminth infection in Pygmy groups indicates that their semi-nomadic lifestyle provides limited, if any, protection against helminths (Cavalli-Sforza, 1986). Rates of helminth infection in Pygmies similar to agricultural Bantu populations may show an increase in sedentism and increasingly agricultural subsistence among Pygmy groups (Ohenjo et al., 2006).

Interestingly, African pastoralist groups have lower STH incidence compared to hunter-gatherer groups, which may be due to their nomadic lifestyle (Teklehaymanot, 2009), although some groups have higher levels of schistosomiasis (Chinwe & Agi, 2014). Fulani pastoralists, who practice cattle husbandry, are required to be in close proximity to freshwater sources which could increase their risk of exposure to schistosoma larvae or cercariae. A recent study found that 66.4% of 593 Fulani pastoralist herdsmen tested positive for infection with *Schistosoma haematobium* (Chinwe & Agi, 2014). The level of helminthiasis and schistosomiasis may depend on periodic movements of pastoralists between seasonally contaminated environments and encampment locations (Teklehaymanot, 2009).

Epidemiological surveys of viral infections in indigenous Africans are rare and never exhaustive, due to a combination of population isolation, small group size, government resources, and NGO/donor interest (Africa, 2006; Froment, 2014). Most observations of viral disease are based on reports from survivors of disease and small children, and indicate that yellow fever, hepatitis C, foamy viruses, and hemorrhagic fevers all have a varying degree of presence in indigenous African communities (Froment, 2014; Hotez, 2014). Around 10% of Bakola pygmies in Cameroon were found to have antibodies to the Ebola virus, which had not been seen in the region before and may indicate infection with a weakly pathogenic strain that could be protective against more severe forms of Ebola hemorrhagic fever (Calattini et al., 2007). Marburg viruses, which are clinically indistinguishable from Ebola, are endemic to woodlands of equatorial Africa, and cases have been reported in Kenya, the Democratic Republic of Congo, Angola, and Uganda

(Adjemian et al., 2011; Bausch et al., 2003; Towner et al., 2006). Fatality rates vary greatly depending on the outbreak, ranging from 23-90% (Leffel & Reed, 2004). Like Ebola, Marburg viruses[4] are spread through human to human transmission and consumption of undercooked African bushmeat containing the virus (MacNeil & Rollin, 2012). Both Ebola and Marburg have no specific antiviral treatment, and although there have been vaccine candidates for Ebola, there are none that are yet approved for clinical use in humans (Henao-Restrepo et al., 2017; Schandock et al., 2017).

Effects of parasitemia and HIV on the gut microbiome

Investigations of the microbiome, immune function, and helminths have focused on how parasite presence in the gut may stimulate mammalian immune response (Allen & Maizels, 2011; Elliott & Weinstock, 2012; Hooper, Littman, & Macpherson, 2012; Littman & Pamer, 2011; Macpherson & Harris, 2004). Many enteric protozoa have a fecal-oral route of transmission, introducing them into the same gastrointestinal space as gut microbiota. It is uncertain if helminths directly modulate the immune system or indirectly affect it through changes in microbiota (Bancroft, Hayes, & Grencis, 2012; Cox et al., 2014; Leung & Loke, 2013).

Gut bacterial families *Elusimicrobiaceae* and *Ruminococcaceae* were important predictors of asymptomatic infection with *Entamoeba* parasites in rural Cameroonian fishers and hunter-gatherer groups (Morton et al., 2015). Gut bacterial composition

[4] The fruit bat *Rousettus aegypti* is the natural reservoir of Marburg viruses and can transmit the virus to humans; the putative reservoir of Ebola is also thought to be a fruit bat (Paweska et al., 2012).

overall produced models that were highly predictive of *Entamoeba* infection (79% accuracy) (Morton et al., 2015). A separate study on diarrheal illness in Bangladeshi children found that levels of *Prevotella copri* were significantly associated with *Entamoeba* infection (Gilchrist et al., 2016). *P. copri* has also been associated with autoimmune disease and severe inflammation (Scher et al., 2013), indicating that *Entamoeba* may trigger an inflammatory response in the distal colon. Mice treated with segmented filamentous bacteria (SFB) were protected from experimental infection with *E. histolytica* and showed increased levels of intestinal neutrophils and IL-23 (Burgess et al., 2014). This immune response suggested that commensal bacteria may alter responsiveness of cytokine induction to inflammatory challenges induced by *Entamoeba* infection.

Higher abundances of *Proteobacteria,* lower abundances of *Verrumicrobia* and *Bacteroidetes,* and higher ratios of *Firmicutes* to *Bacteroidetes* are thought to have a non-specific protective effect against *Cryptosporidium* infection (Chappell et al., 2016). The indole producing bacteria *Escherichia coli* CFT073*, Bacillus spp.*, and *Clostridium spp*. had relative higher abundance in the guts of uninfected humans. Indole could directly inhibit the parasite, stimulate anti-inflammatory pathways, or improve host intestinal barrier function (Chappell et al., 2016; Jin et al., 2014; Shimada et al., 2013). A study on a population from the southern Côte d'Ivoire found that infection with *Giardia duodenalis* was positively associated with increases in *Bifidobacterium* (Iebba et al., 2016), and *in vitro* models have shown that *G. duodenalis* survival was significantly inhibited when *Lactobacillus johsonii* La1 was introduced at the trophozoite phase of

infection (Humen et al., 2005; Pérez et al., 2001). Gut dysbiosis associated with *Giardia* infection has been demonstrated to persist as long as six years after the infection is resolved (Hanevik et al., 2014; Wensaas et al., 2012). *Bifidobacterium*, along with *Streptococcus*, were found in higher proportions in individuals infected with *Plasmodium falciparum* (Yooseph et al., 2015). *Plasmodium falciparum* causes blood and liver stage malaria infections, which are usually treated with antimalarials. These antimalarials can induce metabolic dysregulation, which could in turn affect microbiota composition via modulation of the metabolome or immunity. *Bifidobacterium* is normally considered a beneficial microbe of the gut, so its presence in infected individuals prompts questions about its role in mediating immune responses and gut equilibrium.

Gastrointestinal helminth infection with *Necator americanus, Ascaris sp.,* and *Trichuris trichiura* has been linked to increased alpha (within-individual) and beta (between-person) diversity (See Table 1-1) (Giacomin et al., 2015, 2016; Zaiss et al., 2015). Rural Malaysians with *T. trichiura* or *A. lumbricoides* infections have a greater number of observed taxa, overall species richness, and higher amounts of *Paraprevotellaceae* than non-infected Malaysian controls (Lee et al., 2014). In a large cross-sectional study of rural populations from Indonesia and Liberia, twelve bacterial taxa were significantly associated with helminth (*A. lumbricoides, T. trichiura*, and *N. americanus)* infection (Rosa et al., 2018). Most notably, *Olsenella* was significantly associated with infection and with reduced gut inflammation, and decreased in abundance after infection. Functional categorization of gene content from Indonesian and Liberian helminth infected individuals pointed towards a role for arachidonic acid metabolism as a

precursor for pro-inflammatory leukotrienes that could reduce helminth survival *in vivo* (Rosa et al., 2018).

Infection with *Strongyloides stercoralis* has been associated with an increase in families *Leuconostocaceae, Paraprevotellaceae, Ruminococcaceae,* genus *Peptococcus*, several strains of *Clostridia* implicated in maintaining gut homeostasis, an increased amount of amino acids, and lower levels of short chain fatty acids (SCFAs) (Jenkins et al., 2018). *Leuconostocaceae* has been experimentally shown to release inflammatory Th1 cytokines IL-12 and IFN-γ, producing an antimicrobial immune response (Jenkins et al., 2018); however, other studies on ascarids, whipworms, and hookworms have shown a decrease in *Leuconostocaceae*, warranting further investigation (Jenkins et al., 2017). Bacterial-derived SCFAs exhibit anti-inflammatory properties, and these molecules may play a role in therapeutic effects of helminths used for the treatment of inflammatory disorders. Celiac disease patients that were experimentally treated with *N. americanus* showed increases in SCFAs during infection (Zaiss et al., 2015). However, patients infected with *S. stercoralis* showed significantly decreased amounts of acetate, butyrate, and propionate (Jenkins et al., 2018), indicating that parasite induced changed in bacterial-derived SCFA production may vary by parasite. Differences in cohort sizes and demographics between the *S. stercoralis* and *N. americanus* studies could also contribute to contrasting results.

Some of the regulatory mechanisms influencing helminth-microbiome-host immune functioning in mice co-infected with *Trichinella spiralis* and murine norovirus were recently described (Osborne et al., 2014). Helminth infection upregulated Th-2 cytokines

(IL-4, IL-13) and transcription of STAT6-dependent macrophages. Antiviral immunity was inhibited at the cost of suppressing the helminth parasites, and there were no changes in the bacterial microbiota that correlated with the immune response. Taken together, these results indicated that helminths may have mechanisms that allow direct regulation of antiviral immunity and the host immune system independent of gut microbiome changes (Osborne et al., 2014).

Human immunodeficiency virus (HIV) has a substantial role in disturbing gut homeostasis. HIV infection causes a severe T-cell depletion in the human gut, which in turn compromises gut epithelial integrity and causes microbial translocation from the gastrointestinal tract into the circulatory system, which results in further immune activation and increased HIV replication and pathogenesis (Brenchley et al., 2006; Mudd & Brenchley, 2016). Thus, there is an expectation that HIV infection, development of AIDS, and antiretroviral therapy (ART) will affect structural, metabolic, immunological and functional aspects of the gut microbiome. Prior work on HIV and the enteric bacterial microbiome indicated that chronic HIV infection in western populations being treated with ART correlated with changes in bacterial beta diversity, depletion of *Bacteroides*, and increases in *Prevotella* (Dillon et al., 2014; Dinh et al., 2015; Catherine A. Lozupone et al., 2013, 2014; McHardy et al., 2013; Vázquez-Castellanos et al., 2015; Vujkovic-Cvijin et al., 2013). However, observational studies have proved contradictory, with some studies finding differences in the gut microbiomes of HIV positive individuals in comparison to HIV negative individuals (Goedert, 2016; McHardy et al., 2013; Mutlu et al., 2014; Williams, Landay, & Presti, 2016) and others finding no differences.

In Africa, a study of anorectal swabs from 130 homosexual men in Nigeria failed to find differences between HIV uninfected and infected (ART-naïve) men; however, there was an increased abundance of *Firmicutes, Campylobacter*, and lower *Prevotella* in the gut microbiomes of individuals who were HIV positive and had undergone ART treatment (Nowak et al., 2017). Monaco et al. (2016) were also unable to identify differences in the gut microbiomes of HIV positive and negative Ugandan men, but found that alterations in the bacterial microbiome occurred when their sample population was stratified by CD 4 T-cell numbers, with lower counts linking to increased *Enterobacteriaceae* and enteric adenovirus expansion. Higher abundances of *Prevotella* in these populations, as compared to *Bacteroides* in sub-Saharan Africans could mask increased abundances of *Prevotella* associated with HIV in western cohorts. This research indicates that therapies designed on western populations targeting bacterial microbiomes as an intermediary for treating HIV and AIDS may be ineffective at treating individuals from non-industrialized countries. Additionally, existing microbiome studies on cohorts of sub-Saharan HIV and AIDS positive individuals have focused almost exclusively on men, despite the fact that women and young girls make up a disproportionate amount of new HIV infections (>44% higher than men for girls aged 15-24) (UNAIDS, 2017).

Chapter 3. Population structure of human gut bacteria in a diverse cohort from rural Tanzania and Botswana

The contents of this chapter have been previously published as:

> Hansen, M.E.B. [†], Rubel, M.A. [†], Bailey, A.G., Ranciaro, A., Thompson, S.R., Campbell, M.C., Beggs, W., Dave, J.R., Mokone, G.G., Mpoloka, S.W., Nyambo, T., Abnet, C., Chanock, S.J., Bushman, F.D., Tishkoff, S.A. (2019). Population structure of human gut bacteria in a diverse cohort from rural Tanzania and Botswana. *Genome Biology* 2019 **20**:16. doi: 10.1186/s13059-018-1616-9
>
> [†]Contributed equally

3.1. Abstract

Gut microbiota from individuals in rural, non-industrialized societies differ from those in individuals from industrialized societies. Here, we use 16S rRNA sequencing to survey the gut bacteria of seven non-industrialized populations from Tanzania and Botswana. These include populations practicing traditional hunter-gatherer, pastoralist, and agropastoralist subsistence lifestyles and a comparative urban cohort from the greater Philadelphia region. We find that bacterial diversity per individual and within-population phylogenetic dissimilarity differs between Botswanan and Tanzanian populations, with Tanzania generally having higher diversity per individual and lower dissimilarity between individuals. Among subsistence groups, the gut bacteria of hunter-gatherers are phylogenetically distinct from those of both agropastoralists and pastoralists, but those of agropastoralists and pastoralists were not significantly different from each other. Nearly half of the Bantu-speaking agropastoralists from Botswana have gut bacteria that are very similar to the Philadelphian cohort. Based on imputed metagenomic content, U.S. samples have a relative enrichment of genes found in pathways for degradation of several

common industrial pollutants. Within two African populations, we find evidence that bacterial composition correlates with the genetic relatedness between individuals. Across the cohort, similarity in bacterial presence/absence compositions between people increases with both geographic proximity and genetic relatedness, while abundance weighted bacterial composition varies more significantly with geographic proximity than with genetic relatedness.

3.2. Background

Gut microbiota have been shown to be affected by numerous factors, including host diet, medications, pets, socioeconomic status, environment of residence, and chance acquisition of lineages (Azad et al., 2013; Blaser, Bork, Fraser, Knight, & Wang, 2013; Chen et al., 2012; Chong et al., 2015; Consortium, 2012; Eckburg et al., 2005; Falony et al., 2016; Gill et al., 2006; Harrison & Taren, 2017; Hoffmann et al., 2013; Song, Lauber, et al., 2013; Turnbaugh et al., 2006; Wang et al., 2016; Wu et al., 2011; Zhernakova et al., 2016). While temporary changes in diet have been shown to cause circumscribed shifts in gut bacterial composition, the dominant bacterial composition in healthy adults remains relatively stable and is influenced by long-term diet (Muegge et al., 2011; Turnbaugh, Bäckhed, Fulton, & Gordon, 2008; Turnbaugh, Ridaura, et al., 2009; Wu et al., 2011). Plant and animal domestication during the Neolithic period (~ 10 kya), and the shift from hunter-gatherer subsistence patterns to pastoralist and agriculturalist practices, constituted a major change in diet (Mira, Pushker, & Rodríguez-Valera, 2006). Numerous contemporary, rural African populations continue to practice traditional subsistence lifestyles, including pastoralism, hunting and gathering, and small-scale agropastoralism.

Examining their microbiome composition and function can inform host-microbiota dynamics in the absence of the impact of industrialization and widespread antibiotic use.

Several cross-population studies have compared the gut microbiome of urban-industrial societies with those of traditional hunter-gatherer or agricultural societies. The latter two types of populations consume foods that are relatively lower in sugars, fats, and animal protein and are relatively higher in fiber (Arumugam et al., 2011; Koren et al., 2013; Wu et al., 2011). The gut bacteria of urban-industrialized populations often have high abundances of *Bacteroides*, while the gut bacteria from traditional hunter-gatherer or agropastoral societies have higher abundances of *Prevotella* (Clemente et al., 2012; De Filippo et al., 2010; Gomez et al., 2016; Martínez et al., 2015; Obregon-Tito et al., 2015; Schnorr et al., 2014; Yatsunenko et al., 2012). Whether these trends are due to the types or quantities of foods consumed, cultural or social practices, geographic, genetic, or other factors is unclear. Although there have been several studies of microbiome diversity within African populations (De Filippo et al., 2010; Gomez et al., 2016; Morton et al., 2015; Rampelli et al., 2015; Schnorr et al., 2014; Smits et al., 2017; Yatsunenko et al., 2012), the range of gut microbiome compositions among African populations with diverse subsistence practices remains largely unknown.

Here, we present a comparison of gut microbiota from rural populations in Tanzania ($N = 60$), Botswana ($N = 54$), and individuals living in an urban U.S. city (Philadelphia, PA) ($N = 12$) (Figure 3.1, Table 3-1) (Minot et al., 2013; Ni et al., 2017; Wu et al., 2011). The African populations are composed of multiple ethnic groups practicing varying

degrees of hunting and gathering, agropastoralism, and pastoralism. “Pastoralists” are defined here as any population whose diet and economy are centered on cattle herding. We term populations whose diet and economy are centered around small-scale subsistence farming as “agropastoralists,” as every farming village we sampled also raised cattle or small livestock. Any population that derives most of its food from foraged plants and/or hunted game animals are termed “hunter-gatherers.”

The four Tanzanian populations sampled are (1) the Khoesan click-speaking Hadza, who are savannah hunter-gatherers; (2) the Khoesan click-speaking Sandawe, who are former savannah hunter-gatherers that adopted agropastoral practices over a hundred years ago; (3) the Nilo-Saharan-speaking Maasai, who are semi-nomadic cattle herders; and (4) the Afroasiatic-speaking Burunge, who are agropastoralists. The three Botswanan groups sampled are (1) the Khoesan click-speaking San, who are hunter-gatherers of the Kalahari desert that have recently adopted some agropastoralist practices (Hitchcock, 2002; Ikeya, 2001); (2) the Niger-Kordofanian Bantu-speaking Herero, who are Kalahari pastoralists; and (3) several groups of Niger-Kordofanian Bantu-speaking agropastoralists, hereafter referred to as “Bantu agropastoralists.” The U.S. cohort is mainly composed of individuals who self-identified as “White,” with one self-identified “African American.”

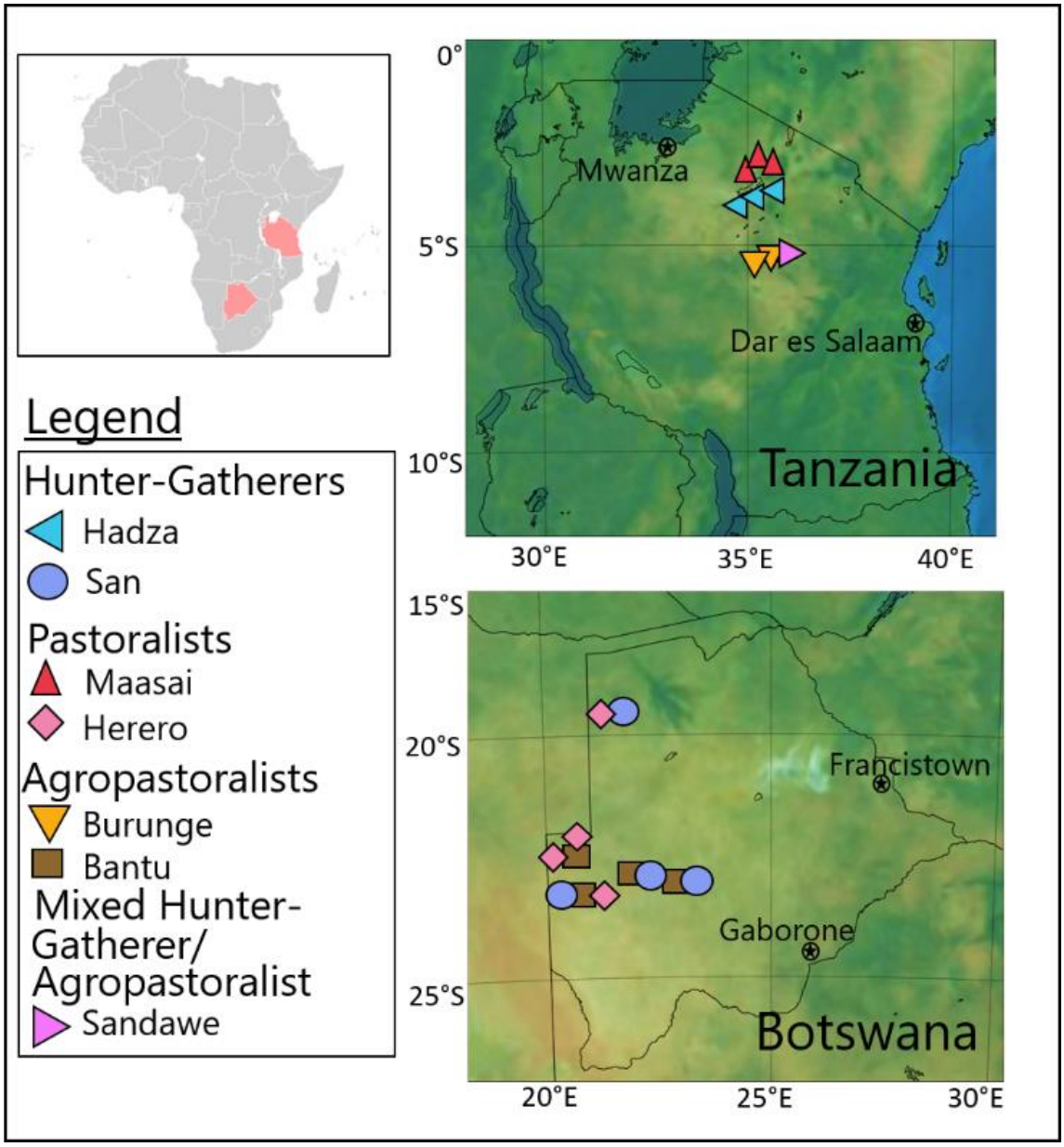


Figure 3-1. Map of the sampled population groups in Tanzania and Botswana.

Country	Population	Subsistence	Number Total	Number Female	Number Male	Age Ave	Age Min	Age Max
Tanzania	Burunge	Agropastoralist	11	10	1	48	22	70
	Sandawe	Agropastoralist	12	10	2	47.2	33	61
	Maasai	Pastoralist	12	6	6	39.5	24	68
	Hadza	Hunting & Gathering	25	10	15	44.2	19	90
		subtotals:	60	36	24	44.5	19	90
Botswana	Bantu	Agropastoralist	26	19	7	49.8	24	92
	Herero	Pastoralist	8	7	1	44.5	19	77
	San	Hunting & Gathering	20	15	5	28.0	18	42
		subtotals:	54	41	13	40.9	18	92
U.S.	Philadelphia	Industrial agropastoralist	12	4	8	26.2	22	33
		Totals:	126	81	45	41.2	18	90

Table 3-1. Botswana and Tanzania cohort metadata per population group
This metadata includes country, population name, subsistence practice, number of individuals, and age range.

3.3. Results

3.3.1. Data overview

DNA was extracted from stool samples, and the 16S rRNA gene V1-V2 segments were amplified and sequenced in all 126 participants. Sequences were aggregated at 97% identity, yielding 18,915 operational taxonomic units (OTUs). Seventeen thousand eight

hundred seventy OTUs mapped to one of 191 bacterial taxa in the Greengenes classification database (D. McDonald et al., 2012), 1,044 OTUs were unassigned, and one OTU could only be mapped at the taxonomic resolution of Kingdom (Bacteria). The mean population abundance of unassigned reads was less than 0.15%, and we removed the 1,044 unassigned OTUs and single Kingdom (Bacteria) OTU from further analysis (Supp. Figure 3-1). Compared to the U.S. samples, the African samples have a larger relative abundance of OTUs that were not confidently assigned to a known taxa (Figure S1A). The four Tanzanian populations have the largest number of unassigned OTUs per individual (Supp. Figure 3-1B), while the Sandawe have a larger number of total unassigned reads per individual compared to any other population (Figure S1B). Collector's curves showing the rate that new OTUs are detected as sample size is increased were calculated for OTUs with abundance > 0.01% and averaged per population (Supp. Figure 3-2). These curves show that increasing our sample size would only marginally increase OTU counts. On average, the Sandawe have the highest number of OTUs, while the U.S. have the lowest (Figure 3-2).

3.3.2. Abundance of *Prevotellaceae* varies within and between African populations

Bacteroidales (phylum *Bacteroidetes*) and *Clostridiales* (phylum *Firmicutes*) are the two most common orders of bacteria in nearly every individual (Figure 3-2A), as expected for the human gut microbiome (Lozupone et al., 2012). The relative proportions of *Bacteroidales* and *Clostridiales* varies by individual and by population (Figure 3-2C). Comparing each population against the rest of the cohort and considering just the two

taxa *Bacteroidales* and *Clostridiales*, we find that the Hadza have a significantly higher proportion of *Bacteroidales* (Mann-Whitney-Wilcoxon (MWW) test, *p-value* 6.3×10^{-4}), the U.S. have a significantly lower proportion of *Bacteroidales* (MWW test, *p-value* 0.020), whereas no other population had a significantly different proportion of *Bacteroidales* (smallest MWW test *p-value* is 0.27).

Prevotellaceae is the most common bacterial family among the Africans in this cohort, being the most abundant family in 70.2% of Africans as well as having the largest mean abundance per population in every African population (Figure 3-2B).
Higher *Prevotellaceae* abundance has been previously associated with infection by the globally endemic gastrointestinal parasite *Entamoeba* in central African rainforest hunter-gatherers (Kirk et al., 2015; Morton et al., 2015). Fecal DNA was screened for *E. histolytica* but this parasite was not detected in our samples, demonstrating that the high *Prevotellaceae* abundances are not due to *E. histolytica* infection in our samples. *Ruminococcaceae* is the second most common bacterial family in the African cohort, being the most abundant bacteria in 14% of Africans.

Bacteroidaceae is the most common bacterial family among the US cohort, being the most abundant family in 50% of U.S. samples and having the largest population mean abundance. *Ruminococcaceae* is the second most common bacteria in the US cohort, being the most abundant bacteria in 25% of the US samples and having the second largest mean abundance.

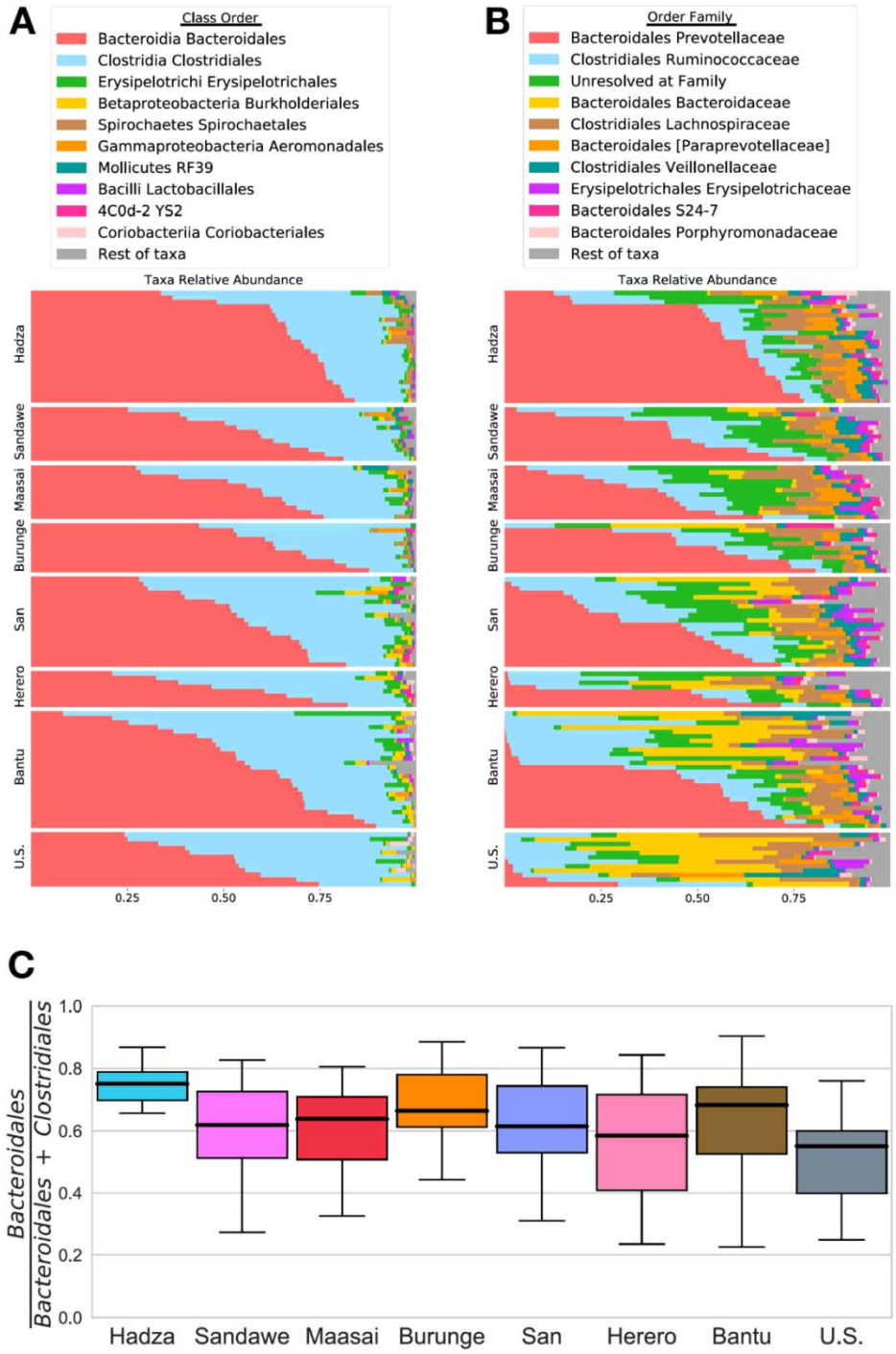


Figure 3-2. Relative abundance per individual for the ten most common taxa
The relative abundance per individual for the ten most common taxa, shown for the bacterial taxonomic rank of A) Order and B) Family. C) The population distribution of the relative proportion of *Bacteroidales* per total of *Bacteroidales* and *Clostridiales*.

Considerable variation in taxa abundances exists within African samples. In particular, 23 Africans have abundances of *Prevotellaceae* as low as what we find in the U.S. samples (within one standard deviation of the mean US, an abundance of 12.4%). Among the African samples with such low *Prevotellaceae* abundance, 19 are from Botswana and of those, 12 are from the Bantu population. Fifty-two Africans were tested by quantitative PCR for absolute 16S rRNA copy numbers per gram of stool, including eight Bantu from Botswana. Of these eight Bantu, six individuals were in the low *Prevotellaceae* Bantu subset, and this group had the lowest average 16S rRNA copy number per gram of stool among any of the African groups (Supp. Figure 3-3, Supp. File 2, Supp. Table 3-1, "Bantu_2"). We tested whether age, sex, host BMI, sampling latitude, or sampling longitude distinguished these 12 individuals from the other Bantu, but none were statistically significant (Wilcoxon rank sum tests, smallest *p-value* is 0.41). Finally, we note that for 12 African individuals, their most abundant bacterial family is not *Prevotellaceae*, *Ruminococcaceae*, or *Bacteroidaceae*, and in ten of these samples the most abundant bacterial family is unresolved.

3.3.3. Bacterial diversity per individual is higher in Tanzania than in Botswana

The African populations varied in gut microbial α-diversity (bacterial diversity, or bacterial richness and evenness, within each individual), as quantified with the Shannon diversity index (Figure 3-3A). The U.S. cohort had the least bacterial diversity, while the Sandawe had the highest, similar to previous comparisons of industrialized populations versus hunter-gatherers (Clemente et al., 2012; Gomez et al., 2016; Morton et al., 2015;

Obregon-Tito et al., 2015; Schnorr et al., 2014) and small-scale agropastoralists (De Filippo et al., 2010; Yatsunenko et al., 2012). These trends are not impacted by the rarefaction of OTU counts to 5000 per individual, as evidenced by the high correlation in Shannon diversity index with and without rarefaction (Spearman's rho correlation 0.998, Supp. Figure 3-4). The α-diversity was not significantly correlated with the absolute 16S rRNA gene copy number ($R2 = -0.011$, $p\text{-}value = 0.51$) (Supp. Figure 3-3E). However, we do find that the absolute 16S rRNA gene copy number is significantly higher in the Tanzanians than in Botswanans (MWW test, FDR q value = 0.023) (Supp. Figure 3-3D).

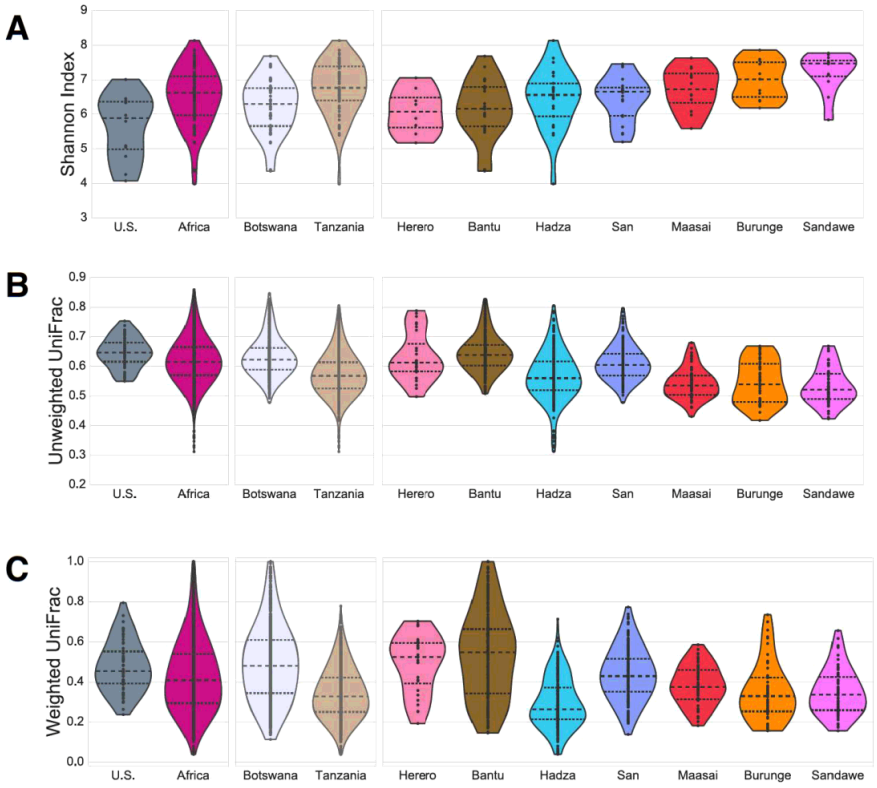


Figure 3-3. Within group mean a and b diversity.
A) Shannon index B) Unweighted UniFrac distance distribution within each group. C) Weighted UniFrac distribution within each group.

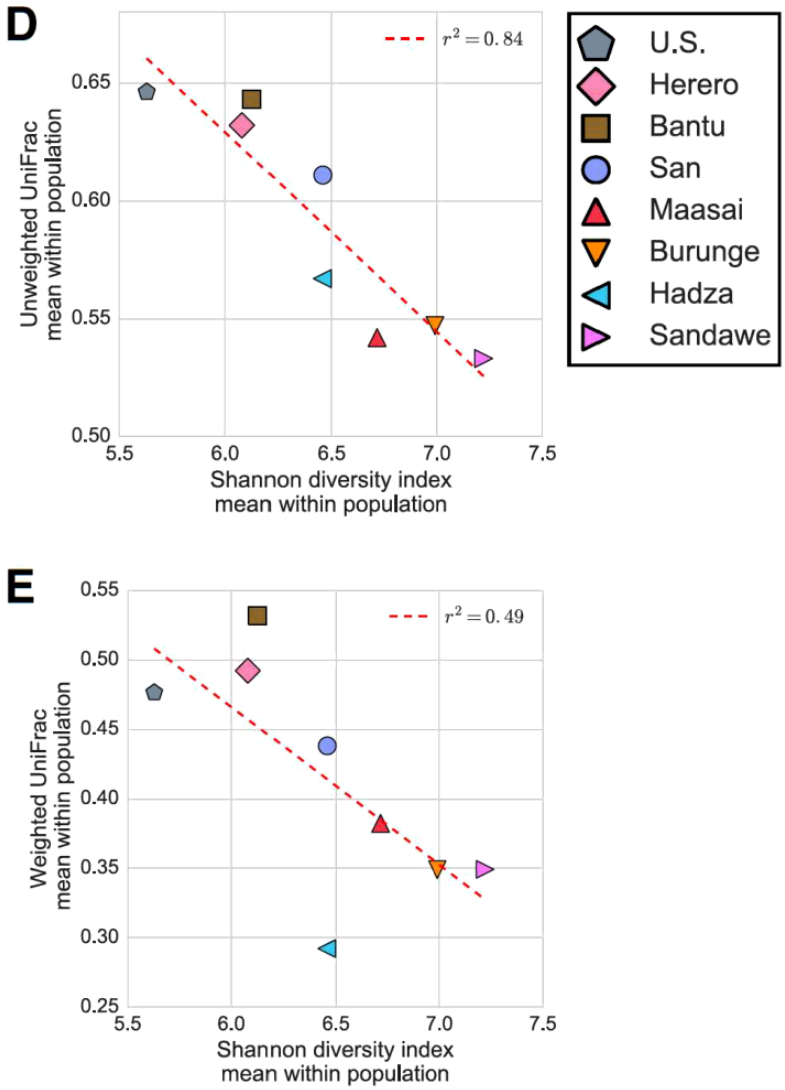


Figure 3-3 (cont.). Within group mean a and b diversity The within-population mean Shannon diversity versus unweighted and weighted UniFrac distances are shown in D) and E), respectively.

3.3.4. Between host bacterial diversity is correlated with within-host bacterial diversity

The populations also varied in the within-population β-diversity (microbiota compositional dissimilarity between hosts), quantified by the UniFrac distance. The UniFrac distance is the fraction of the phylogenetic tree not shared between two samples, where the phylogeny of all taxa found in a bacterial community is estimated based on

ribosomal RNA sequence similarity. Smaller values indicate greater sharing of the microbial phylogenetic tree among hosts within a population, which may reflect greater homogeneity in environmental factors (e.g., diet, cultural practices, shared geographic location). The Tanzanian Sandawe have the lowest within-population β-diversity, while the Botswanan Bantu and US have the largest within-population β-diversity (Figure 3-3B,C).

We find a significant negative correlation between mean population α- diversity and mean within-population β-diversity for unweighted UniFrac distances (Figure 3-3D,E) (unweighted UniFrac β-diversity: linear regression $R^2 = 0.84$, $p\text{-}value = 1.47 \times 10^{-3}$, and Kendall Tau correlation -0.79, $p\text{-}value = 6.5 \times 10^{-3}$; weighted UniFrac β-diversity: linear regression $R^2 = 0.49$, $p\text{-}value = 0.052$, and Kendall Tau correlation -0.43, $p\text{-}value = 0.14$). When individual pairs are restricted to the same sampling location for the within-population UniFrac calculation, the trend across Tanzanian populations is no longer evident, though the differences between Tanzania and Botswana remain (Supp. Figure 3-5). The correlation between α- diversity and β-diversity also holds when counts are rarefied to 5000 reads per individual (Supp. Figure 3-6), which accords with the high degree of correlation in UniFrac distances with and without rarefaction (Spearman's rho correlation of 0.965 and 0.999 for unweighted and weighted UniFrac, respectively, Supp. Figures 3-7 and 3-8). Additionally, the Bray-Curtis dissimilarity metric for β -diversity yields similar results as the weighted UniFrac distance (see Supp. Figures 3-9 and 3-10). Thus, the correlation between α- and β-diversity does not appear to be an artifact of

choice of UniFrac as a β-diversity measure, uneven sampling location diversity, or uneven sequencing depth across individuals.

3.3.5. Gut bacteria composition is more distinct between countries than between subsistence practices

The gut bacterial compositional differences between populations were quantified by the mean UniFrac distance between all pairs of individuals taken from between-population pairs. The between-population bacterial phylogenetic distances were larger between the U.S. cohort and each African population than between any two African populations (Supp. Figure 3-11, Additional File 2, Supp. Tables 3-2, 3-3). The largest unweighted UniFrac distance within Africa was between the Botswanan Bantu and Tanzanian Hadza, which represented 92% of the average distance between the U.S. and African populations. The largest weighted UniFrac distance within Africa was between the Bantu and the Herero in Botswana, which is nearly 84% of the mean distance average between the US and African populations.

The degree of compositional difference between two groups was assessed with PERMANOVA (Anderson, 2001) tests of UniFrac distances, which measures the significance of the between-group variation to within-group variation (pseudo F-statistic) by permutation of group assignment. If two groups have identical distributions of bacterial composition, then the pseudo F-statistic will be ~ 1, with larger values corresponding to greater difference in composition between the two groups. As shown in Figure 3-4, among pairs of countries, the U.S. and Tanzania have the largest pseudo F-statistic. The pseudo F-statistic for Tanzania and Botswana is as large, or nearly as large,

as the pseudo F-statistic between the U.S. and Botswana (Figure 3-4A,D, Additional File 2, Supp. Table 3.4-3.7). This result demonstrates that the bacterial compositional variation between two regional, rural, African cohorts can be of similar magnitude as the compositional variation between an urban/industrialized cohort and a rural African cohort.

Among the three pairs of African subsistence groups, the hunter-gatherers have significantly different compositions from both the agropastoralists and the pastoralists, while the agropastoralists and pastoralists are not significantly different from each other (Figure 3-4B,E, Additional File 2, Supp. Table 3-8, 3-9). Comparing the magnitudes of difference between subsistence groups and geographic groups, we therefore find that the bacterial compositional difference between Tanzania and Botswana (a geographic grouping) is larger than between any of the African subsistence groups (both unweighted and weighted UniFrac F-statistics). From this observation, we infer that the gut bacteria are phylogenetically more distinct between groups defined by region (country) than by subsistence practice.

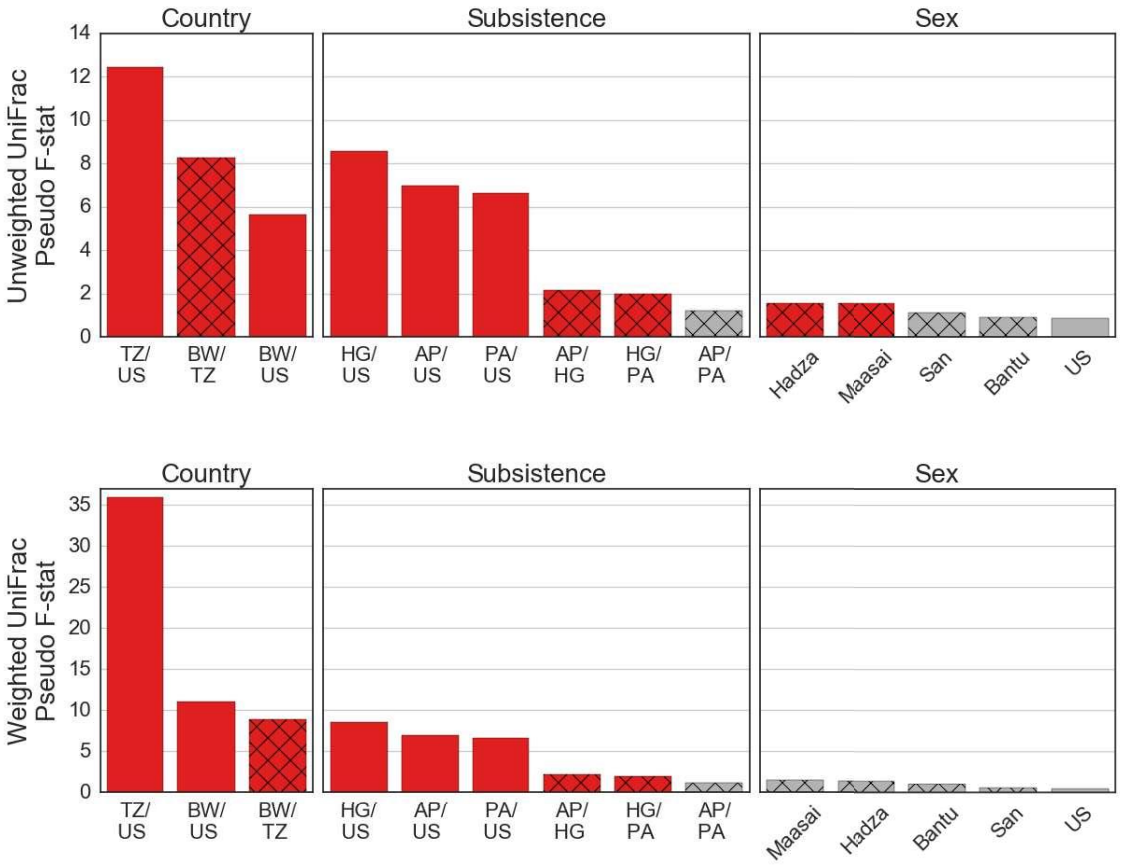


Figure 3-4. PERMANOVA tests of the phylogenetic difference between pairs of groups, based on unweighted UniFrac and weighted UniFrac. Shown are groups defined by country of origin (panels A, D), subsistence practice (panels B, E), and sex (panels C, F). Unweighted UniFrac panels are A-C and weighted UniFrac panels are D-F. The subsistence practices are abbreviated as U.S. = western (Philadelphian), HG = hunter-gatherers (Hadza, San), AP = agropastoralists (Bantu agropastoralists, Burunge, Sandawe), and PA = pastoralists (Herero, Maasai). Bar in red denote pairs where the F-statistic p-value is < 0.05.

3.3.6. Gut bacterial composition is significantly different between males and females in the Maasai and Hadza

Four populations were tested for differences between sex in bacterial α-diversity (Shannon diversity index, minimum of five individuals per sex for MWW test) (Hadza, Maasai, San, and Bantu), and none showed a significant difference (all MWW test *p-values* > 0.17). In addition, five populations were tested for a significant β-diversity distance between sexes using PERMANOVA (UniFrac distances, minimum of four

individuals per sex) (Hadza, Maasai, San, Bantu, and U.S.). The Hadza and Maasai had significantly larger unweighted UniFrac distances between sexes than expected by chance (PERMANOVA p-value < 0.05) (Figure 3-4C, Additional File 2, Supp. Table 3-10), while no population had a significantly elevated weighted UniFrac distance between sexes (all PERMANOVA p-values > 0.2, Figure 3-4F, Additional File 2, Supp. Table 3-11). Thus, there appears to be elevated phylogenetic differences between sexes in the Hadza and Maasai in terms of presence or absence of bacterial OTUs but not in terms of OTUs weighted by their abundance. Although we find a significant difference between sexes for these two populations, larger sample sizes will be needed to identify the factors causing these differences.

3.3.7. Gut bacteria compositions of individuals from the U.S. are more similar to Botswanans than to Tanzanians

Using principal coordinate analysis (PCoA), we find that the similarities in overall bacterial OTU composition among individuals are strongly correlated with the abundances of three common bacterial families, *Prevotellaceae*, *Bacteroidaceae*, and *Ruminococcaceae* (Spearman's rho correlation with PCo1 *p-value*s are $1.0 \times 10{-}40$, $1.0 \times 10{-}18$, and $3.3 \times 10{-}12$, respectively, and Spearman's rho correlation with PCo2 *p-value*s are $5.0 \times 10{-}7$, $6.0 \times 10{-}2$, and $1.1 \times 10{-}20$, respectively) (Figure 3-5, Supp Figure 3-12). The first principal coordinate (45% of variance) is most strongly associated with *Prevotellaceae* abundance while the second principal coordinate (12% of variance) is most strongly associated with *Ruminococcaceae* abundance.

The per population distribution of weighted UniFrac distances between Africans and the U.S. cohort shows that the Botswanan gut bacteria are phylogenetically more similar to the U.S. gut bacteria than the Tanzanian gut bacteria (Figure 3-5D). This observation is consistent with the PERMANOVA results and the observation that there are more people with low *Prevotellaceae* abundance and high *Bacteroidaceae* abundance in Botswana than in Tanzania.

The 13 Bantu with high *Prevotellaceae* abundance similar to other Africans were markedly more different from the US cohort based on both weighted and unweighted bacterial composition as well as within-individual bacterial diversity (weighted UniFrac PERMANOVA test *p-value* = $2.0 \times 10{-}5$; unweighted UniFrac PERMANOVA test *p-value* = $2.0 \times 10{-}5$; MWW test on Shannon diversity *p-value* = 0.014). By contrast, the 12 Bantu with low *Prevotellaceae* abundance, similar to the US cohort, were not statistically different from the U.S. samples based on bacterial abundance (weighted UniFrac PERMANOVA test, *p-value* = 0.12). However, they were different based on unweighted bacterial composition and within-individual bacterial diversity (unweighted UniFrac PERMANOVA test, *p-value* = $3.0 \times 10{-}4$ and MWW test on Shannon diversities, *p-value* = 0.028, respectively). Thus, the similarities between these Bantu and US individuals are driven by common bacteria.

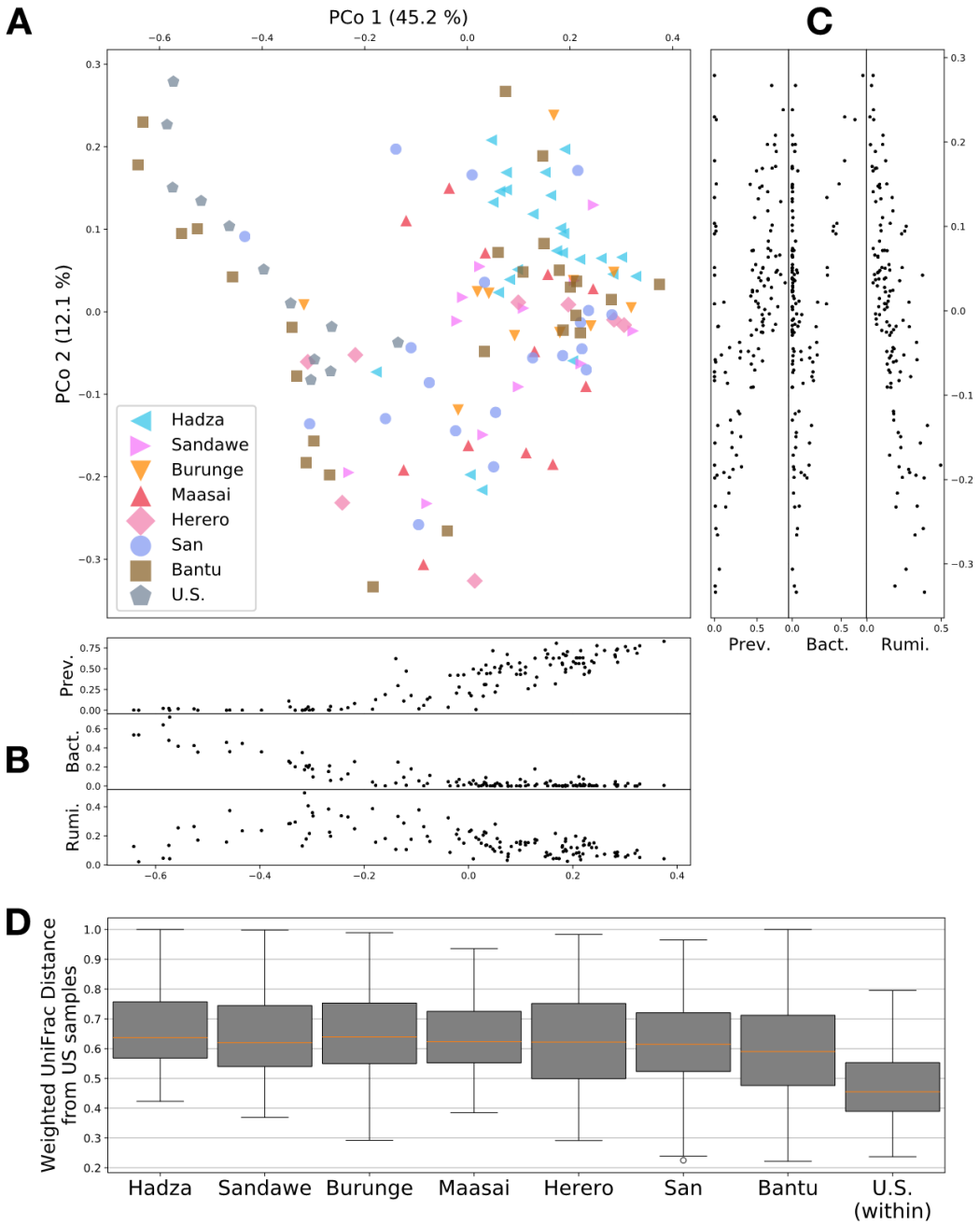


Figure 3-5. Principle Coordinate Analysis (PCoA) for weighted UniFrac distances. The first two principle coordinates for all individuals in the study, where marker shape and color denote the population of origin. Sidebar panels B and C show the abundances of *Prevotellaceae* (Prev.), *Bacteroidaceae* (Bact.), and *Ruminococcaceae* (Rumi.) aligned to the first two principal coordinates. D) Box-and-whisker distributions between each African population and the U.S. samples, over all pairs of individuals.

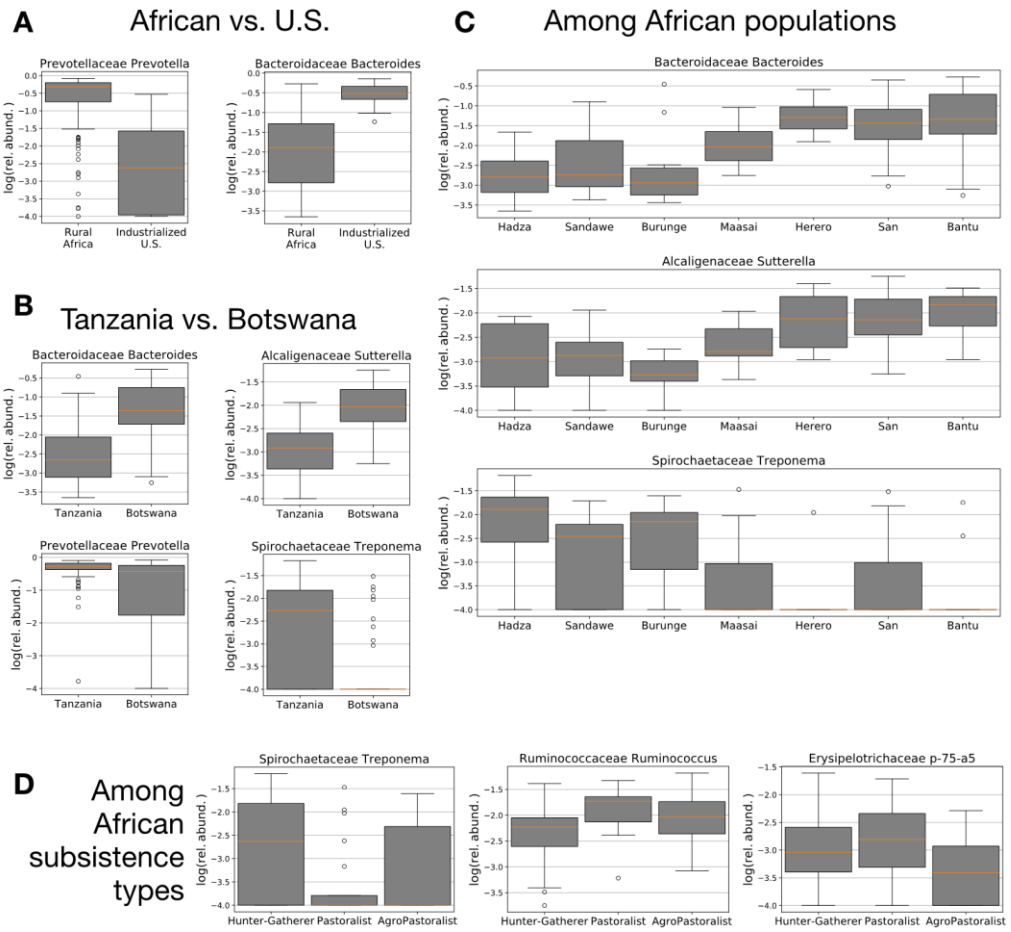
3.3.8. Observations of differentially abundant bacterial families among populations, subsistence groups, age, and sex

The Analysis of Composition of Microbiomes (ANCOM) method (Mandal et al., 2015) was used to test for significantly differentially abundant bacteria among groups defined by country, population, subsistence lifestyle, and sex. We found that two bacterial genera (out of $N\text{genus} = 48$) vary significantly between Africa and the U.S., *Bacteroides* and *Prevotella* (Figure 3-6A), both of which also varied significantly between Tanzania and Botswana (Figure 3-6B). We also observed that *Bacteroides* is one of several bacteria that are differentially abundant among the seven African populations (Figure 3-6C). We note that 43.3% of all *Bacteroides* reads, and 35.2% of African *Bacteroides* reads, came from a single OTU (denovo36).

Among the three African subsistence categories, three genera varied significantly (Figure 3-6D): *p-75-a5*, *Ruminococcus*, and *Treponema*. *p-75-a5* has previously been found in fecal samples from healthy children from Bangladesh (Lin et al., 2013) as well as in pre-weened calves (Malmuthuge, Griebel, & Guan, 2014). *Ruminococcus* is also found in both human and ruminant fecal samples (Russell & Rychlik, 2001). The fact that both *p-75-a5* and *Ruminococcus* bacteria have the highest abundance in pastoralists may be the result of close interaction between humans and livestock. The third genera that varied significantly among subsistence groups, *Treponema*, was most abundant in hunter-gatherers and agropastoralists, and has been previously associated with hunting and gathering and small-scale agropastoral populations with diets high in fiber (Gomez et al., 2016; Obregon-Tito et al., 2015; Schnorr et al., 2014). Within the African cohort, no taxa

were found to vary significantly between sexes nor among three age classes (18-39, 40-59, 60+).

Figure 3-6. Box-and-whisker plots of relative abundances distributions per group for the taxa that varied significantly among groups by the ANCOM tests. Individuals are grouped



by A) traditional or industrial lifestyle, B) country of origin, C) population, and D) traditional subsistence strategy. Relative abundances were log-transformed to account for the wide range of values (e.g., the means between Africans and the U.S. in A are ~2 orders of magnitude different).

3.3.9. Imputed metagenomes show functional differences between populations and countries

The functional variation among populations was predicted using the metagenomic imputation method Phylogenetic Investigation of the Communities by Reconstruction of

Unobserved States (PICRUSt) (Langille et al., 2013). For every individual and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, PICRUSt estimates the total gene count within that pathway (normalized to a relative abundance per pathway). Individuals were then grouped by population, subsistence, country, and continent, and statistical tests were computed on differences in the distribution of pathway abundances. One hundred forty-six KEGG pathways were significantly differentially enriched between the U.S. and African cohorts, and 148 KEGG pathways were significantly differently enriched between Botswana and Tanzania (White's nonparametric *t* test, FDR < 0.1) (Additional File 2, Supp. Table 3-12:3-15). The pathway abundances of the Botswanan cohort were almost always intermediate between those of the Tanzanian and the U.S. cohorts. The pathway relative abundance difference between Tanzania and Botswana was highly correlated with the pathway relative abundance difference between Africa and the U.S. (Spearman's rho correlation 0.51, *p-value* < 10−22; Supp. Figure 3-13A). We infer that the regional differences in bacterial abundances may lead to regional differences in functional pathway abundances, depending upon the accuracy of gene content imputation. For example, we find that the degradation pathway of the pesticide dichlorodiphenyltrichloroethane, commonly known as "DDT," is enriched in Botswanan samples but not in Tanzanian or U.S. samples (Supp. Figure 3-13B). No KEGG pathways varied significantly among African subsistence groups (ANOVA, FDR > 0.1). Twenty-six KEGG pathways were significantly differentially enriched both between Africa and the U.S. and between Tanzania and Botswana and also have absolute relative differences > 15% between continent and between country (Additional File 2, Supp. Table 3-12 and Supp. Figure 3-13A, shown in red). These are the pathways with the most extreme

regional differences in pathway enrichment. Among these, five involve antibiotic biosynthesis or resistance, six involve the degradation of industrial xenobiotic compounds, nine involve digestion, and three involve cell recognition or cell-cell signaling. The frequencies for antibiotic resistance pathways and the xenobiotic degradation are highest in the U.S. and lowest in Tanzania, while the biosynthesis of the antibiotic ubiquinone has highest pathway frequency in Tanzania and lowest in the U.S.

3.3.10. Gut bacterial alpha-diversity is higher in people with low BMI

Previous research in humans and mice have observed correlations between lower α-diversity and prevalence of obesity (Haro et al., 2016; Turnbaugh, Hamady, et al., 2009; Turnbaugh et al., 2006). Across all individuals ($N = 126$), we find that the α-diversity was significantly negatively correlated with the age- and sex-regressed BMI values (Kendall tau (KT) correlation -0.21, *p-value* $= 6.2 \times 10{-}4$) (Additional File 2, Supp. Table 3-17), indicating lower bacterial diversity in individuals with higher BMI. The correlation remained significant among just the African individuals ($N = 114$, KT correlation -0.18, *p-value* $= 5.1 \times 10{-}3$), although it was not significant within any single population (each population *p-value* > 0.05). There is, however, a significant negative correlation within the pastoralists (KT correlation -0.42, *p-value* $= 0.0094$) and within the agropastoralists (KT correlation -0.22, *p-value* $= 0.023$). In addition, we tested for correlation (Kendall tau) between the regressed BMI values and the abundance of each bacteria at the taxonomic rank of genus with at least a 0.1% relative abundance in at least one population (*Ngenus* = 56) (Additional File 2, Supp. Table 3-18. 3-19). We observed

that 11 bacteria were significantly correlated with BMI (FDR < 0.01), the most significant of which are *Treponema* and *Anaerovibrio*.

To test whether the between-population differences in mean BMI (regressed on age and sex) drive the correlation between BMI and α-diversity over all individuals, we constructed "population re-centered" residuals by subtracting the population mean BMI from each individual's BMI, according to their population of origin. The resulting correlation between the "population re-centered" BMI residuals and the α-diversity was not statistically significant (KT p-value > 0.1 over all samples and over African samples only, see Additional File 2, Supp. Table 3-17). Similarly, none of the bacterial taxa at the rank of genus are significantly correlated with the population re-centered BMI residuals (FDR > 0.5). From this observation we conclude that the significance of the correlation between BMI and α-diversity is due to between-population differences. Thus, we cannot rule out that other host environmental or cultural covariates affecting BMI may be associated with bacterial diversity and abundance.

3.3.11. Bacterial compositional similarity increases with geographic proximity and inter-individual relatedness

We investigated the differences in gut bacteria based on geographic distance and the degree of host genetic relatedness. A subset of 97 people was densely genotyped using the Illumina 5M SNP array, with at least eight individuals from each African population, allowing for the estimation of their genetic relatedness. To test whether genetic relatedness had any impact on the distribution of bacteria within a population, we calculated the correlation between host genetic relatedness and bacterial UniFrac distance

among all pairs of individuals within each population. Genetic relatedness is quantified by the estimated identity-by-descent fraction, which is the fraction of the genome that is estimated to be identical between two people due to a shared recent common ancestor. Estimation of the identity-by-descent fraction assumes a panmictic population and is, therefore, reasonably suited for use as a within-population relatedness metric. To control for possible differences between sexes, we filtered the pairs of individuals to only include individuals of the same sex.

Only the Hadza have statistically significant correlations between identity-by-descent and both unweighted and weighted UniFrac bacterial distances (Additional File 2, Supp. Table 3-20), while the Maasai have a significant correlation between identity-by-descent and unweighted UniFrac bacterial distance, indicating in both cases that more related individuals have more similar bacterial composition. Considering all tests, the correlation between identity-by-descent and unweighted UniFrac distance is negative in all but one case (unweighted UniFrac among the Herero). The probability that all seven weighted UniFrac correlations are negative by chance is < 0.01 (sign test), while the probability that at least 6 of 7 weighted UniFrac tests are negative by chance is 0.0625 (sign test). Thus, while we detect a statistically significant correlation between host relatedness and bacterial phylogenetic overlap only in the Hadza and the Maasai, there is a general trend for more related individuals to have more similar bacterial composition.

In addition, we examined the joint impact of geography and host relatedness on bacterial composition with a linear analysis of UniFrac distances. We modeled the bacterial phylogenetic distance (UniFrac) between hosts as a linear function of the host genetic

relatedness and the host geographic separation: $U_{ij} \sim D_{ij} + G_{ij}$, where i and j are index individuals, U_{ij} is the bacterial UniFrac distance, D_{ij} is the geographic distance between the sampling sites for the individuals (measured in kilometers), and G_{ij} is the genetic relatedness of individuals. Here, we quantify G_{ij} with the correlation of normalized and centered genotype counts (Speed & Balding, 2015; J. Yang et al., 2010). This relatedness measure is widely used to control for population structure in cohorts drawn from multiple mating populations (e.g., genetic principal components analysis or as the covariance structure of random effects in linear-mixed models of genetic association tests) and, thus, is well suited as a measure of genetic relatedness when considering differences across genetically diverse populations.

The genetic relatedness and geographic distance between sampling sites are highly correlated (Spearman's rho correlation -0.66, $p\text{-}value < 10^{-10}$). We therefore regressed G_{ij} on D_{ij} and used the residuals, G'_{ij}, when fitting the model $U_{ij} \sim D_{ij} + G'_{ij}$ to the observed data using linear least squares. For unweighted UniFrac bacterial distances, the best fit coefficients of D_{ij} and G'_{ij} are both significantly non-zero (T test $p\text{-}value$s < 0.002, Additional File 2, Supp. Table 3-21), indicating that bacterial similarity is greater with closer geographic proximity and closer relatedness. For weighted bacterial UniFrac distances, only the coefficient of the geographic separation is significantly non-zero (T test $p\text{-}value < 0.001$, Additional File 2, Supp. Table 3-21). Although a linear model can only capture the main trends of the complex processes that shape the observed distribution of the gut microbiome, it serves to indicate that bacterial composition varies

with geographic proximity and that the stratification with host relatedness is larger for bacterial presence/absence data than for abundance weighted data.

3.4. Discussion

We surveyed the bacterial composition of fecal samples from rural populations in Tanzania and Botswana and a comparative population from Philadelphia in the U.S. Among the rural Tanzanian and Botswanan populations, there are population level differences in bacterial diversity and abundances. We also found correlations between host BMI and both overall microbial diversity (less diverse microbiota were correlated with higher BMI) and the abundances of specific taxa. Host genetic similarity is correlated with more similar bacterial composition within the Hadza and Maasai populations. When comparing across African populations, we find genetic relatedness is correlated with presence/absence of gut bacteria, even when accounting for geographic separation.

The bacterial community diversity we observe between rural African populations is comparable to that observed by Gomez et al. (2016) between two groups from the Central African Republic, the BaAka hunter-gatherers and a neighboring group of Bantu-speaking agriculturalists. The unweighted UniFrac distances between the BaAka and the neighboring Bantu is nearly 74% of the mean distance between the U.S. and the African cohort, while for weighted UniFrac this ratio is nearly 70%. The African populations in our cohort are slightly more phylogenetically diverse based on unweighted UniFrac

distances (83% of distance between U.S. and Africa) and are comparable in terms of weighted UniFrac distances (68% of distance between U.S. and Africa).

Given the difficulty of directly comparing microbiome studies that use different amplicons for OTU measurements, we contextualize our results with taxa-level meta-analysis from Smits et al. (2017), which identified four bacterial families and one bacterial phylum that primarily associate with traditional (*Prevotellaceae, Spirochaetaceae, Succinivibrionaceae*) or industrialized (*Bacteroidaceae, Verrucomicrobia*) populations (Supp. Figures 3-14, 3-15; Additional File 2, Supp. Table 3-22). Additionally, three of these five taxa (*Succinivibrionaceae, Spirochaetaceae,* and *Prevotellaceae*) were highly variable with season. With the inclusion of our study cohorts, this modified meta-analysis has bacterial compositional data from 26 populations in 17 countries (34 cohorts). For the U.S. cohort used in this study, the mean abundances of the five taxa were within a standard deviation of the mean values for one or more U.S. cohorts in the Human Microbiome Project (Supp. Figures 3-14, 3-15; Additional File 2, Supp. Table 3-22), indicating that it is not an outlier compared to prior studies.

The relative abundance of *Prevotellaceae* in the Hadza from our study was ~58%, which is nearly ten-fold higher than the 8% relative abundance of *Prevotellaceae* found in the Hadza by Schnorr and colleagues (2014) (Schnorr et al., 2014), although it is within a standard deviation of the relative abundance reported by Smits et al. (2017) (~38%) (Supp. Figures 3-14, 3-15; Supp. Table 3-22). Prior studies of the Hadza (Smits et al., 2017) and the Hutterites from the U.S. (Davenport et al., 2014) indicate that seasonally

volatile gut bacterial taxa correlate with seasonally available food. The Hadza in our study were sampled mid-October through early November, which is the late dry season and the beginning of the wet season, when there is an average rainfall of 57 mm (The World Bank Group, n.d.). Schnorr and colleagues sampled from the Hadza population in the rainy season during January (Schnorr et al., 2014), when average rainfall is ~146.6 mm (The World Bank Group, n.d.). The increased abundance of *Prevotellaceae* in the Hadza in our study is concordant with seasonal variation of this taxa reported in Smits et al. (2017). As with any bacterial taxon and study population, differences in *Prevotellaceae* abundance between microbiome studies of the Hadza could be affected by use of different protocols, reagents, and primers.

Fluctuations in short-term diet could also explain some of the variability seen between microbiome studies (David et al., 2014) in the Hadza and our other sampled populations. Although we unable to obtain individual or population level dietary information for our research participants, we conducted a nutritional literature review to provide a qualitative assessment of contemporary diet in the traditional populations presented in this study (see Supp. Table 3-23 and Methods for extended dietary information). Given the dissimilarity of food types between industrialized and traditional populations, the compositional similarity between the Bantu and U.S. is noteworthy and may be reflective of individual nutritive changes in the Bantu from Botswana and a shift from traditional to industrialized diets. It is clear from Figure 3-2 that there is heterogeneity in bacterial abundance profiles within the Bantu, where roughly half the population has gut bacteria similar to the other African groups, and the other half has gut bacteria more similar to the

U.S. cohort. We could not identify any host factors (age, sex, BMI, location) that significantly distinguish these two groups of Bantu. If we are observing a population undergoing changes in life styles that impact gut bacteria, then the changes in gut bacteria are not uniform across the population. Future work aimed at pairing longitudinal gut microbiome research with individual and population level dietary surveys would be informative for determining the extent to which shifts in subsistence and diet affect microbial changes.

Bacteroides has been used to distinguish between developing (low *Bacteroides* abundance) and industrialized (high *Bacteroides* abundance) populations (Clemente et al., 2012; De Filippo et al., 2010; Gomez et al., 2016; Martínez et al., 2015; Morton et al., 2015; Obregon-Tito et al., 2015; Schnorr et al., 2014; Yatsunenko et al., 2012) and is significantly variable across the African populations, with generally higher abundances in Botswanans than in Tanzanians. We find that the gut bacterial composition of the U.S. population is closer that of the Botswanan populations than to any of the Tanzanian populations. In particular, the U.S. gut bacterial composition was most similar to the Botswana Bantu agropastoralists, and 12 of the Botswana Bantu agropastoralist individuals have gut bacteria that are not significantly different from the U.S. individuals by abundance weighted composition. The U.S. and Botswana Bantu agropastoralists also have the two lowest measures of taxonomic diversity within hosts and two of the highest measures of inter-individual diversity in this cohort. Botswana is more economically developed than Tanzania, reflected in higher yearly per capita gross national income ($15.5 k in Botswana to $1.75 k in Tanzania), and a higher percentage of

Botswanans (57%) than Tanzanians (30%) live in urban areas (Regional Office for Africa, Africa Health Organization, World Health Organization, 2016b, 2016a). None of the populations in our study live in an urban setting; sites are ~ 60 km or more walking distance from the nearest town (see Additional File 2, Supp. Table 3-24 for more details). It is possible that there are country-level differences in pathogens, sanitation, hygiene practices, transportation access, or medical access between Tanzania and Botswana that impact gut microbiome composition in rural areas.

The Hadza, San, and Sandawe are three current or former hunter-gatherer populations in various stages of settlement or transition from their ancestral subsistence lifestyle. The Sandawe settled into villages and adopted small-scale agropastoral practices in the mid-1800s (Newman, 1970). The Sandawe have the greatest bacterial α-diversity in the cohort, which may be related to their genetic admixture with neighboring populations and/or their mix of subsistence practices. Varying subsistence strategies could plausibly increase gut bacterial diversity 1) neutrally, through the introduction of a wide array of microbes due to a varied life style and diet, or 2) selectively, due to bacterial community adaptation to varying environments. The high Shannon diversity values in the Sandawe are consistent with the "Intermediate Disturbance Hypothesis" which proposes diversity of bacteria is maximized under conditions of fluctuating environments (e.g., diet in this case) (Jiang & Patel, 2008; Kadmon & Benjamini, 2006).

The Hadza are unique in this cohort in that they still largely practice hunting and gathering (Marlowe, 2010). Their gut bacteria are outliers in several respects: A) they

have the highest abundance of *Prevotellaceae* and *Spirochaetaceae* (Clemente et al., 2012; Gomez et al., 2016; Morton et al., 2015; Obregon-Tito et al., 2015; Schnorr et al., 2014), in particular, the genus *Treponema* within family *Spirochaetaceae*, which is a common constituent of hunter-gatherer gut microbiomes (Clemente et al., 2012; Gomez et al., 2016; Morton et al., 2015; Obregon-Tito et al., 2015; Schnorr et al., 2014) and a catabolizer of fibrous plant materials (cellulose and xylans) that form a large component of Hadza diets (Schnorr et al., 2014), B) the Hadza are outliers in unweighted UniFrac PCoA, indicating that their bacteria, in terms of presence/absence, are phylogenetically the most dissimilar to other African populations, C) they and the Maasai are the only two populations (out of five tested) with a significantly distinct microbiome between sexes, D) the Hadza common gut bacteria are phylogenetically more homogenous across the population relative to all other populations in this study, and E) the within-population variation in their gut bacteria is correlated with the relatedness among individuals, where more related individuals tend to have more similar bacterial composition (both presence/absence and abundance weighted). The bacterial phylogenetic differences between sexes in the Hadza that we observe corroborates a previous finding of sex differences in the Hadza gut microbiome (Schnorr et al., 2014), and may be partly attributable to sexual division of labor and differential food intake (Schnorr et al., 2014). Hadza men and women have different activity patterns, where men are highly mobile foragers with access to honey and game meat, while Hadza women forage for local materials and may engage in more frequent "snacking" on fiber-enriched foods than men (Berbesque, Marlowe, & Crittenden, 2011; Pontzer et al., 2012).

The Maasai of Tanzania and the Herero of Botswana are two cattle herding peoples that live in close proximity to domesticated animals and have a heavy dairy component to their diet (International Livestock Centre for Africa, 1991). However, the Maasai and Herero gut bacteria are not more similar to each other than to those of other neighboring populations in their respective countries. The Maasai, like the Hadza, have a significant distinction in bacterial communities between sexes. Maasai men are in charge of supervising and herding cattle (International Livestock Centre for Africa, 1991; Nestel, 1985) whereas women traditionally manage the household, oversee milk production from animals and milk distribution (or sale), and supervise small livestock (goats, sheep) (International Livestock Centre for Africa, 1991; Nestel, 1985). The separation of labor and the time away from home spent by men while tending cattle (Christensen et al., 2012) could affect the types and quantities of food that men eat compared to women.

Across the seven African populations, we find a significant negative correlation between α and β-diversity, which corroborates a trend that has been previously observed between pairs of Western and non-Western populations (Clemente et al., 2012; De Filippo et al., 2010; Gomez et al., 2016; Martínez et al., 2015; Obregon-Tito et al., 2015; Schnorr et al., 2014; Yatsunenko et al., 2012). Several implications follow from this general trend: first, the negative correlation between α and β-diversity exists among a set of non-Western populations practicing largely traditional subsistence lifestyles, demonstrating that the correlation is not entirely a Western versus non-Western phenomenon; second, the correlation is not associated with the particular subsistence lifestyle; and third, the correlation is more significant for unweighted β-diversity than for

abundance weighted β-diversity. These three points, and the fact that the correlation is negative, are consistent with a neutral, diffusion-limited process accounting for most phylogenetic differences in gut microbiome communities between the African populations in our study. This does not argue against selection acting on specific bacteria according to their niche role, only that selection on broad subsistence type does not appear to determine the overall phylogenetic distance between populations (Etienne, 2005).

The contribution of host genetics to gut microbiome composition remains an open question, with studies finding evidence for heritability of relative bacterial abundances or specific taxa (Bonder, Kurilshikov, et al., 2016; Etienne, 2005; Goodrich et al., 2016, 2014; Qin et al., 2010; Turpin et al., 2016; Xie et al., 2016), and alternately, estimating that host genetics explain only a minor percentage of microbiome variation (Rothschild et al., 2018). We do not know whether the observed correlations between bacterial composition and host genetic similarity that we find in the within-population analysis of the Hadza and Maasai, or in the joint analysis of geographic and relatedness across all population, is tracking differences in specific genetic factors that mediate interactions with commensal microbiota (e.g., inflammation response or mucin production genes), or the tendency for closely related individuals to live and/or work in the same places and hence have a greater degree of shared environment compared to unrelated individuals. The bacterial compositional differences seen between countries, between populations within a country, and the significant dependence on the geographic distance between individuals in a linear model, underscore the importance of physical separation on the

distribution of gut bacteria among population groups. Longitudinal studies may be may be required to understand whether these correlations are plausibly due to bacterial dynamics within a population, while much larger cohorts are required for adequately powered statistical tests of whether these correlations are due to heritable host genetic factors.

The functional differences of the predicted metagenomic content of the gut microbiomes supports the hypothesis that there are both country-level and population-level differences in the distribution of functional pathways among the gut bacteria. We find that most imputed KEGG pathways that are more enriched in the USA compared to the two African countries as a whole are also more enriched in Botswana than in Tanzania. KEGG pathways with this enrichment pattern include categories that relate to the degradation of industrial compounds and by-products, such as bisphenol, xylene, DDT, and styrene. This pattern possibly reflects selection for increasing the abundance of bacteria that can degrade or metabolize environmental xenobiotic compounds.

The imputed bisphenol degradation pathway also has highest frequency in the USA, followed by Botswana, then Tanzania. Bisphenol is a common industrial organic compound used in many plastics and epoxies. The sampled African populations live far from industrial centers and arguably have less contact with plastics and industrial by-products compared to the U.S. individuals; consequently, the frequency pattern of the bisphenol degradation pathway could indicate that the presences of bisphenol is influencing the composition of the U.S. gut microbiome. A similar argument applies to

the higher frequencies of imputed styrene degradation and xylene degradation pathways in the U.S. compared to Botswana and Tanzania.

Within Africa, we find that Botswana has a higher frequency of these industrial compound degradation pathways compared to Tanzania, including imputed DDT degradation pathways. Interestingly, Botswana, but not Tanzania, is one of nine countries worldwide that uses indoor residence spraying of traditional structures for control of malaria-carrying mosquitos (Chihanga et al., 2016; World Health Organization & Global Malaria Programme, 2017). These results suggest potential metagenomic adaptation to increased exposure to industrial compounds in western populations, and to DDT in Botswanans.

There are caveats to interpretation of PICRUSt results; we do not know with certainty what variables explain the differences in imputed functional enrichment. Additionally, imputed gene content from reference strains may not adequately capture the gene content in strains that have diverged due to, for example, horizontal gene transfer and selection (e.g., antibacterial resistance). Shotgun sequencing of the gut bacteria will be required to directly verify the metagenomic functional differences observed here and to investigate potentially novel bacterial strains found in these Africa populations. The U.S. population sampled here is the only population from an urban city in our study, which we may reasonably expect to contain more industrial pollutants in the general environment than in the environment of any of the populations we sampled in Africa. Consequently, it would be of interest to sample populations from Botswana and Tanzania that reside in major urban centers where there is more exposure to industrial pollutants, to see if their gut

bacteria are enriched for functions more similar to what we see in the U.S. population with regard to industrial by-product degradation and xenobiotic metabolism.

3.5. Conclusions

The genetic and cultural diversity of Africans extends to the taxonomic diversity of their gut microbiomes. The gut bacteria in Botswana are relatively more similar to the U.S., and a subset of traditional farmers has gut bacteria nearly indistinguishable from that in the U.S. cohort. Correspondingly, the phylogenetic diversity between rural African populations can be as large as the differences we find between traditional and urban populations. In general, the regional phylogenetic distinction between Botswana and Tanzania exceed the distinction found between subsistence lifestyles. The factors causing a shift towards Western microbiome compositions remain unknown but appear to have a regional component that is not entirely due to differences in agricultural, pastoral, or hunting-gathering subsistence modes.

3.6. Methods

3.6.1. Sampled populations

Ethnic groups, language, sample sizes, subsistence classifications, and sampling coordinates of populations are listed in Additional file 2: Table S1A. Written informed consent was obtained from all participants, and ethics/research approval and permits were obtained from the following institutions prior to the start of sample collection: NIMR, COSTECH, and Muhimbili University of Health and Allied Sciences in Dar es Salaam, Tanzania; The University of Botswana and the Ministry of Health in Gaborone,

Botswana; IRB approval from the University of Pennsylvania. Samples were collected from Botswana during the wet season and the start of the dry season (January–April) and from Tanzania at the end of the dry season/start of the wet season (October–March). We recruited 114 adult participants (26 Bantu, 8 Herero, 20 San, 25 Hadza, 12 Sandawe, 12 Maasai, 11 Burunge) who practiced diverse modes of subsistence (such as pastoralism, agropastoralism, hunting and gathering, and mixed hunting and gathering). Demographic information including age, sex, ethnicity, and ancestry was recorded (for further participant details see Additional File 2, Supp. Table 3-1). We provide a dietary literature review of our sampled populations in Additional File 2, Supp. Table 3-23. Basic demographic data and fecal 16S rRNA V1-V2 sequences for healthy Philadelphians (U.S. cohort) were collected during prior studies at the University of Pennsylvania (Minot et al., 2013; Ni et al., 2017; Wu et al., 2011). All fecal and blood samples were extracted, sequenced, and analyzed using the same laboratory and computational pipelines, thereby reducing the impact of batch effects in the cross-population comparisons.

In Tanzania, samples were obtained from the Hadza hunter-gatherers who live in the Arusha and Shinyanga regions surrounding Lake Eyasi, the Maasai pastoralists from the northern Ngorongoro district, and the Burunge agropastoralists and Sandawe former hunter-gatherers who reside near each other in the Kondoa district in Central Tanzania (Figure 3-1). Each of these four ethnic groups has a distinct dietary pattern. Hadza hunter-gatherers rely on local, natural resources that are structured by annual and seasonal changes in rainfall (Smits et al., 2017). Specifically, Hadza diets are dominated by tubers, legumes, berries, baobab fruit, honey, and foraged plant material (Marlowe,

2010; Marlowe & Berbesque, n.d.). The Sandawe are a former hunter-gatherer group that settled in villages and began farming in the 19th century. They primarily subsist on grains, with supplements of tubers and plant material gathered from the bush. Up until the mid-1800s, the Sandawe were a semi-nomadic hunter-gatherer population living in the savannahs of Tanzania. The Sandawe have admixed with neighboring populations of diverse ancestries who migrated into Tanzania within the past 5,000 years (Tishkoff et al., 2009). The Sandawe also adopted the agropastoral subsistence practices of neighboring Bantu-speaking Turu, which comprises the bulk of their caloric intake, though they continue to supplement a small portion of their diet with hunting and gathering (Newman, 1970; Yatsuka, 2016).The Maasai are nomadic cattle, sheep, and goat herders living in the Ngorongoro highlands region. Maasai diets primarily consist of meat, milk, and blood, which are lactose rich and high in fat and cholesterol, though they supplement that diet with maize traded from neighboring groups (Århem, 1989). The Burunge are settled farmers that also keep livestock, with a diet heavily dependent on millet and subsidized by cattle derived dairy and meat.

In Botswana, samples were obtained from western/northwestern regions from San populations who traditionally have practiced hunting and gathering (Naro, Kaukau, Ju|'hoan, !Xoo) and from several agropastoralist populations (Kgalagadi, Tswana, Mophadima) that are classified here as "Bantu" based on their shared language family and broad subsistence practice, and one population, the Herero, who practice a pastoralist lifestyle. The traditional diet of San hunter-gatherers is composed of foraged meat, vegetables, fruits, and nuts, the latter of which contributed the largest percentage of

dietary protein and calories (Lee, 1979; Silberbauer, 1965, 1972; Tanaka, 1980; Valiente-Noailles, 1993). Some San settlements receive a substantial component of their food from government sources (Hitchcock, 2002; Ikeya, 2001). Bantu agropastoralists have diets mainly composed of sorghum, maize, millet, legumes, cucurbits (squash and melons), eggs, and seasonally available fruits in addition to goat, chicken, fish, and cattle meat (Grivetti, 1978). Herero pastoralists have diets based on beef, milk, and milk products with supplements of goats, chickens, garden produce, foraged plants and animals, and bulk grains (especially ground corn) (Pennington, 2002).

3.6.2. Sample collection and storage

Participants produced a fecal sample in a sterile container that was immediately returned to researchers at the field site. A midsection sample of stool was harvested in a 5ml container and immediately frozen in liquid nitrogen. Samples were later aliquoted into smaller 1.5 ml containers on dry ice in a fume hood to maximize storage space. The samples were stored at -80°C before transportation to the U.S. in dry ice, where it was again stored at -80°C until extraction.

3.6.3. Biological sample processing and quantification

16S rRNA gene sequencing and processing for microbiome sequencing

Total DNA from fecal materials was extracted using a PSP Spin Stool DNA Plus Kit (Stratec Molecular) with a modified bead-beating method (Salonen et al., 2010). PCR and extraction blanks were used to control for reagent and environmental contamination, and all extractions were conducted in a laminar flow hood, with equipment and consumables

given UV irradiation for a minimum of 30 minutes prior to use. Eluted DNA was quantified by fluorometry and stored at −20°C. PCR reactions were performed in quadruplicate using the Accuprime system (Invitrogen) and barcoded composite primers with Illumina adapters to amplify the V1-V2 sections of the 16s rDNA genome (see Table S1Y for 16s rDNA gene sequencing metadata and PCR conditions). The resulting 300-320 bp products were pooled and visualized by gel electrophoresis, followed by product purification using 1:1 volume of Agencourt AmPure XP beads (Beckman-Colter). Purified PCR products, including extraction and PCR blanks, had their final concentration determined with Qubit PicoGreen dsDNA BR assays (Invitrogen) and were pooled in equal amounts prior to Illumina Nextera XT library preparation (processed by the manufacturer's protocol). Libraries were multiplexed on the Illumina MiSeq system and sequenced using 2 x 250 bp cycles. Sequence data were deposited under project accession PRJNA395034 in the NCBI Sequence Read Archive; sample details and individual accession numbers are included in the Additional File 2, Supp. Table 3-1.

16S rRNA processing and qPCR

In a separate extraction, total DNA from fecal materials was extracted from samples using a MO BIO PowerSoil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA). Eight samples from each population were included save for the Herero, where only three were available, and the Bantu, where nine were available. No stool samples were available for the U.S. individuals, so they were not included in this analysis. Each fecal sample was individually weighed, with samples ranging from 0.012 to 0.196 g. The samples were then processed according to manufacturer's protocols, and eluted DNA was

quantified by fluorometry and stored at −20°C . Bacterial abundance was quantified by qPCR amplification of the V1V2 region of the 16S rRNA gene, with reactions performed in triplicate (25 μL each), using 1:1000 dilutions of DNA template. For qPCR, equal volumes of purified DNA of all samples were used in this assay. Primer and probe sequences are as follows: BSF8 (Forward) qPCR primer – 5´-AGAGTTTGATCCTGGCTCAG-3´, BSR65/17 (Reverse) qPCR primer– 5´-TCGACTTGCATGTRTTA-3´, Fluorescent dye (5´6 -FAM (Fluorescein)), landing sequence, dark quencher (3' Black Hole Quencher®-1) 5'– /56-FAM/TAA +CA+C ATG +CA+A GT+C GA/3BHQ_1/ - 3'. *A + indicates a locked nucleic acid base. Primers and probes were purchased through Integrated DNA Technologies (IDT).

Prior research has indicated that the differences between 16S qPCR copy numbers produced from the same samples but extracted with both PSP and MoBio kit were statistically negligible (Lauder et al., 2016); thus, the MoBio extracts can serve as an accurate proxy for PSP extracts for 16S qPCR. The Bantu had a mean 16S rRNA gene copy number per gram of stool of $1.51 \times 109 \pm 3.73 \times 108$ SEM (standard error of the mean), the Burunge had $8.06 \times 109 \pm 5.01 \times 109$ SEM, the Hadza had $1.81 \times 109 \pm 4.25 \times 108$ SEM, the Herero had $1.68 \times 109 \pm 4.73 \times 108$ SEM , the San had $1.41 \times 109 \pm 3.27 \times 108$ SEM, the Maasai had $1.81 \times 109 \pm 3.35 \times 108$ SEM, and the Sandawe had $1.87 \times 109 \pm 2.12 \times 108$ SEM .

OTU clustering

Bacterial 16S rRNA reads were analyzed using the Quantitative Insights into Microbial Ecology (QIIME) software package (Caporaso, Kuczynski, et al., 2010). During the

quality-filtering process, reads were removed from the analysis if they did not match golay error-corrected barcode with less than two mismatches, if the read pairs could not be joined with an overlapping sequence of less than 35 bp, if they had a homopolymer sequence (repeated base call) greater than 6 bp, and if they had more than two ambiguous base calls (N's). OTUs were created by single-linkage clustering the reads using Swarm (Mahé, Rognes, Quince, de Vargas, & Dunthorn, 2014) and removing OTUs comprised of only a single or pair of reads. Representative sequences from each OTU were aligned using the PyNAST aligner (Caporaso, Bittinger, et al., 2010), and a phylogenetic tree was inferred using FastTree v. 2.1.3 (Price, Dehal, & Arkin, 2010) after applying the standard Lane mask for 16S sequences (Lane, 1991). As an additional quality control step, all OTUs were tested for correlations between the proportional abundance of the OTU and the post-PCR amplicon concentration of a sample using the method developed by Jervis-Bardy et al. (2015) as implemented in the contam_test program for R (https://github.com/eclarke/eclectic). A negative correlation indicates a potential contaminant: an increasing proportional abundance of that OTU in correlation with lower sample biomass (as implied by lower amplicon concentration) suggests that the increased proportional abundance of that OTU comes in as part of the reagents, and is not truly part of the sample. Correlation significance was assessed using Pearson's rho, and OTUs with a significant negative correlation were considered to be contaminants and removed. Final OTU sequences are listed in Supp. Table 3-26. Taxonomic assignments were generated using the Greengenes 16S database v. 13_8 (Daniel McDonald et al., 2012) (Table S1AA) and OTUs mapping to chloroplast or mitochondrial sequences were removed. All OTUs are denoted by the prefix "denovo" since they are determined without use of

reference sequences. OTU and MRT abundances are measured as the proportion of the total reads per individual.

Host genotyping and genetic relatedness

DNA was extracted from white blood cells using a salting out method (Gentra Puregene) and 97 of the 114 African individuals were genotyped on the Illumina Omni5M Exome array that includes a small number of indels and ~4.5 million SNPs (see Supp. Table 3-1). In collaboration with the Cancer Genomics Research laboratory (CGR) at NIH, array intensity data was clustered and all genotypes were called based on standard operating procedures using the hg19/37 SNP coordinates in the Illumina software GenomeStudio. See Crawford et al. (2017) for further details on this genotype callset. We retained the segregating autosomal biallelic single nucleotide polymorphisms (SNPs) over the 97 individuals, and variants were pruned to be in approximate linkage equilibrium, $r^2_{LD} < 0.1$ using plink (Chang et al., 2015) (plink --indep-pairwise 200 kb 20 0.1), leaving 158,891 SNPs for genetic relatedness estimation. From these sites, we constructed A) the estimated pairwise identity-by-descent fraction among all pairs of individuals from the same population (plink --genome, see Supplemental File 1), and B) a genetic relationship matrix between all pairs of individuals i and j from the standardized genotype vectors using the Genome-wide Complex Trait Analysis (GCTA) software (J. Yang, Lee, Goddard, & Visscher, 2011) (--make-grm-gz) for subsequent analyses (see Supplemental Files 2 and 3).

qPCR for *Entamoeba histolytica*

Primers, probes, and protocols for the qPCR, including methods for generating a recombinant plasmid containing target *E. histolytica* sequence to make a standard curve, were taken from Mejia et al. (2013). qPCRs were run on a QuantStudio7 Flex Real-Time PCR system.

3.6.4. Statistical methods

Diversity and richness measurements

Diversity metrics (α and *b*-diversity) were quantified using all 17,861 taxonomically mapped OTUs using QIIME (Lozupone & Knight, 2005; Catherine A. Lozupone, Hamady, Kelley, & Knight, 2007). QIIME was also used to calculate UniFrac distances, which are an estimate of the fraction of the total branch length over the bacterial phylogenetic tree that is not shared by two bacterial communities (Lozupone & Knight, 2005; Catherine A. Lozupone et al., 2007). Unweighted UniFrac distance is based on the presence/absence of bacteria (Lozupone & Knight, 2005), while weighted UniFrac distance weights the shared branches in the phylogenetic tree by abundance (Lozupone et al., 2007). Species accumulation curves were calculated using the *specaccum* function from the vegan library for R.

Phylogenetic variation among groups

PERMANOVA tests between groups were computed with Python package *scikit-bio* (scikit-bio.org), using 50,000 permutations. The PERMANOVA test statistic is the ratio

of between-group to within-group variance, and is sensitive to whether the mean separation between groups is larger than the mean variance within groups.

Principal Coordinate Analysis

Principle coordinate analysis (PCoA) was computed using the Python package *scikit-bio 0.5.1* (scikit-bio.org).

Analysis of differentially abundant taxa

We used the Analysis of Composition of Microbiomes (ANCOM) method to detect differentially abundant taxa between groups (Mandal et al., 2015), as implemented in the Python *scikit-bio* 0.5.1 package. The ANCOM method accounts for the simplex nature of compositional data, and so does not suffer from spurious negative correlations imposed by the fact that (relative) abundances across all bacteria must sum to one within a given bacterial community. This method tests for taxa that vary significantly among groups more than a significant number of the other taxa. Consequently, if a large number of taxa all vary similarly among the groups, then none of these will show up as significantly varying compared to the other taxa. As such, this is a sensitive test for taxa that vary significantly and in an unusual way compared to the other taxa. For all tests, we used the default "one way ANOVA" base test, with a significance threshold of 0.05, tau parameter 0.99, and theta parameter 0.25, and we used the Holm-Bonferroni multiple testing correction.

For these analyses, we used abundances per mapped genus that have at least 0.1% mean abundance in at least one of the eight populations for the between continent comparison,

and at least 0.1% mean abundance in at least one of the seven African populations for Africa-only comparisons. Since we were testing for difference between groups for each taxa, we rescaled the relative abundance by a constant factor and re-centered the relative abundances by adding a constant for each taxa, such that the rescaled relative abundances span from 1/N to 1, where N is the total number of samples. The ANCOM analysis tests for differences between groups using the logarithm of the rescaled abundances. Since the logarithm cannot handle zero values, the choice of 1/N as the minimum rescaled value avoids this issue. For the between continent comparison, all individuals were used, N=126. For the Africa-only comparison we used N=114. The rescaled relative abundances are given by X = (x-A)/ (B-A), where x is the original relative abundance, A=Min (x)-1/N, and B = Max (x). Note that the rescaling was done separately when using all samples (N=126) or the Africa-only samples (N=114).

Functional metagenomic analysis

Subsampled reads were subjected to closed reference OTU picking against the Green Genes reference taxonomy (Greengenes database, May 2013 version; http://greengenes.lbl.gov) using the pick_closed_reference_otus.py script in QIIME(Caporaso, Kuczynski, et al., 2010) using 97% identity. Metagenomes from bacterial OTUs were imputed with PICRUSt on the online Galaxy interface (http://huttenhower.sph.harvard.edu/galaxy). For each individual and each KEGG pathway (N_{KEGG}=328), PICRUSt calculates the cumulative gene count across all OTUs that overlap the pathway, which are then normalized into a pathway abundance. The data were analyzed statistically by using STAMP v. 2.0.6. In this data set, the highest pathway

frequency values are on the order of 0.1-1%, while the smallest non-zero pathway frequencies are on the order of 10^{-7}-10^{-6} %. The Nearest Sequenced Taxon Index (NSTI), which measures the phylogenetic distance between observed OTU sequences and the reference, has a mean and standard deviation of 0.140 and 0.037 across all samples, with 10th and 90th percentiles 0.095 and 0.186 (Table S1AB). For multiple (>2) populations, ANOVA and Tukey-Kramer post-hoc tests were performed. Two-group comparisons were done with White's non-parametric t-test with two-sided confidence intervals obtained by bootstrapping. Multiple tests were controlled with FDR correction calculated by the Benjamini-Hochberg method.

For a given pathway k, the relative abundance difference between two groups A and B, $R_k(A,B)$, is defined by $R_k(A,B) = \frac{X(A) - X(B)}{(X(A) + X(B))/2}$. Across all pathways *k* we find a significant, positive, correlation between $R_k(Africa,U.S.)$ and $R_k(Tanzania,Botswana)$ (Spearman rank correlation 0.51, p-value < 10^{-22}) (Table S1AC).

PICRUSt relies on the assumption that the bacterial strains in each sample have the same gene content as database strains used for the analysis, which can be inaccurate when strains vary substantially in gene content. However, it does not appear that the above correlation in pathway abundances can be explained by annotation biases alone. For example, if OTUs from Tanzania had a lower mapping rate to known taxa compared to those from Botswana, then this would result in lower abundances across all pathways in Tanzania compared to Botswana; this is not what we observe, as many pathways have a higher abundance in Tanzania than Botswana. The above correlation could only be

explained by OTU annotation biases that A) happen to impact certain pathways more than others, and B) have the same mapping biases between Tanzania and Botswana as between the U.S. and Africa. While we cannot rule this out, it would require several biases to align in direction.

BMI correlation

All of the individuals in our study have BMI measurements, which allowed testing for correlations between α-diversity and BMI. The BMI values were regressed on age and sex, and the residuals were tested for correlation with the Shannon diversity index. The "population re-centered" BMI residuals are computed as follows: the mean BMI residual is computed for each population, and for every individual in this population this value gets subtracted from their BMI residual.

Linear regression

Linear modeling and least squares fitting of UniFrac distances as a function of host geographic separation and host genetic relatedness were computed using the Ordinary Least Squares routines in the python package *statsmodels*).

3.7. Declarations

3.7.1. Acknowledgements and funding

We would like to thank Kyle Bittinger and Jesse Connell for their advice throughout the project, and Stephanie Schnorr for stimulating discussions on the Hadza and gut microbiome differences between sexes. Lisa Mattei and the Penn CHoP Microbiome Core were of great help with the 16S qPCR assay, and we wish to thank Ryan Pinti for

his assistance with conducting the nutritional literature review. Thanks also to the members of the Bushman laboratory for help and suggestions. We also thank all the field assistants for their help during the sample collection. Above all, we thank our African research participants for donating their samples, data, and time to this research.

This research was supported by the Lewis and Clark Fund, the University of Pennsylvania, the Leakey Foundation, an NIH training grant in Parasitology (5T32AI007532-18), and the National Science Foundation (BCS-1540432) to MAR and to SAT (BCS-1317217), and NIH awards (DP1 ES022577-04, 1R01DK104339-01 and 1R01GM113657-01) to SAT. MEBH was supported by an NIH postdoctoral training fellowship (T32ES019851-02) through the Center of Excellence in Environmental Toxicology (CEET) at the University of Pennsylvania.

3.7.2. Availability of data and materials

The OTU count data and anonymized participant information is available in Supplementary Material. Tanzanian and Botswanan bacterial 16S sequence data can be found at the Sequence Read Archive (SRA), accession number PRJNA395034. The host genetic data is stored at the NIH database of genotypes and phenotypes, dbGaP (www.ncbi.nlm.nih.gov/gap), accession number phs001396.v1.p1 (see Additional File 2, Supp. Table 3-1 for the dbGaP sample identifications for those individuals involved in the present study) (S. Tishkoff, 2017). Data for the healthy U.S. cohort were collected from prior studies at the University of Pennsylvania (Minot et al., 2013; Ni et al., 2017; Wu et al., 2011).

3.7.3. Authors' contributions

The study design was by SAT, FBD, MEBH, and MAR. African fecal samples were collected by AR, MCC, and SRT with the coordination of SAT, SWM, GM, MB, and TN. US fecal samples were collected by FDB. Blood samples for genotyping were collected by AR, SRT, and MCC. Fecal 16S rRNA sequencing was conducted by MAR with support from JRD, and in coordination with FDB. Blood samples were processed by WRB and genotyping analysis was done by CA and SJC. MEBH and MAR performed data analyses, with assistance from AGB and JRD. The manuscript was written by MEBH and MAR and was edited by SAT in consultation with FDB. All authors have read and approved the manuscript.

3.7.4. Ethics approval and consent to participate

All experiments were approved by the University of Pennsylvania Institutional Review Board (IRB # 807981). Ethical approval and permits were also obtained from the following institutions prior to sample collection: The Commission for Science and Technology and National Institute for Medical Research in Dar es Salaam and the Ministry of Health in the Republic of Botswana. Written, informed consent was obtained from each participant.

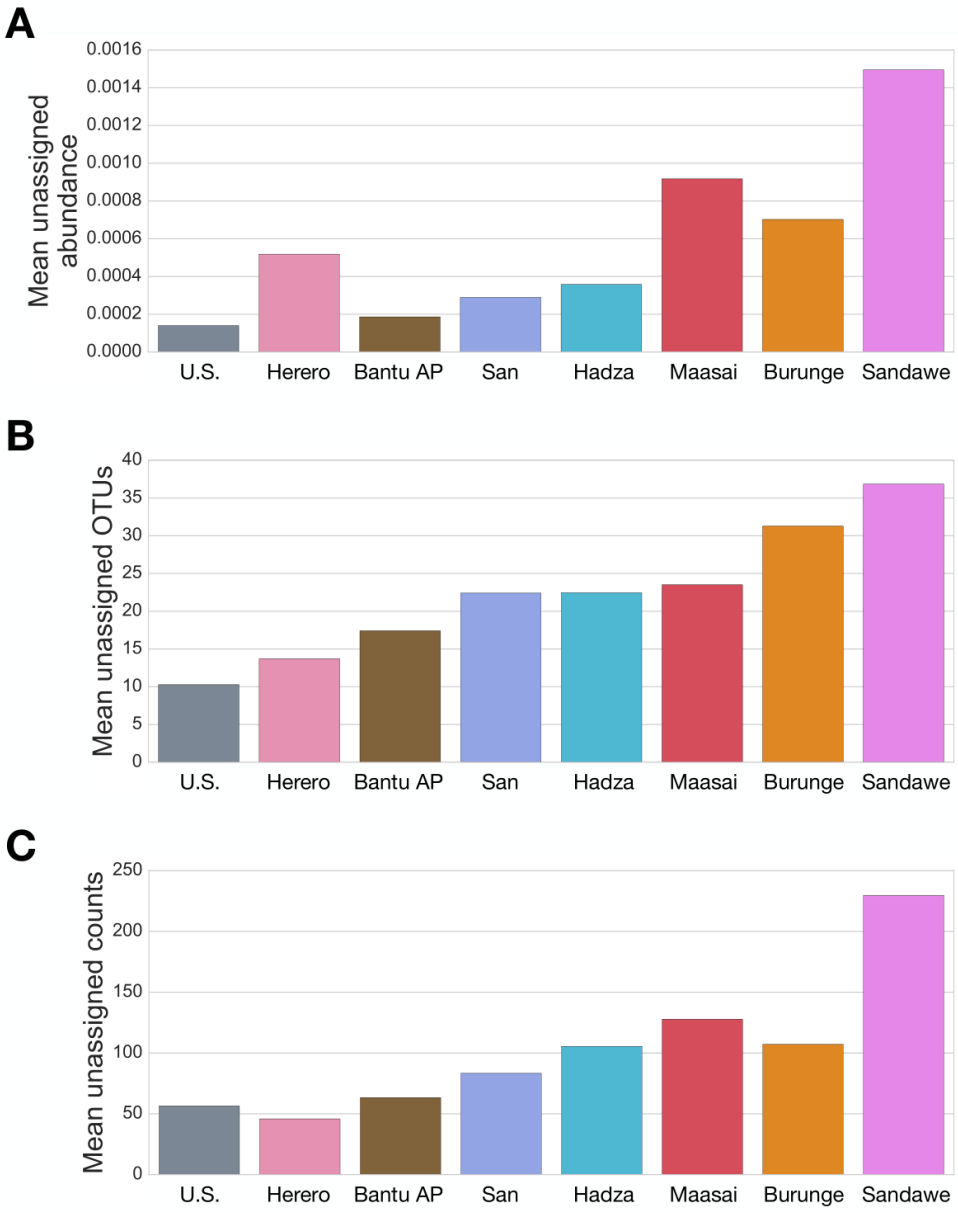
All experimental methods were in accordance with Helsinki Declaration.

3.7.5. Consent for publication and competing interests

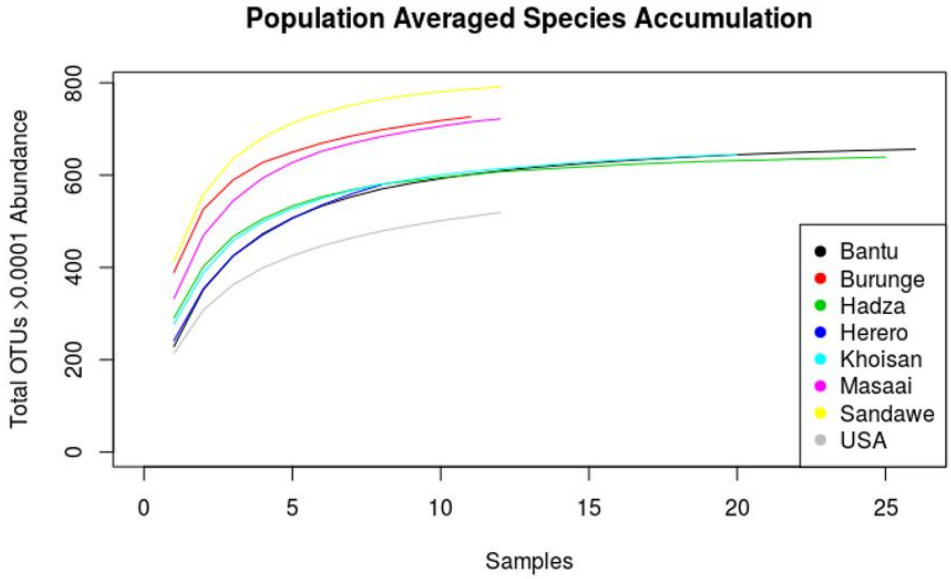
All participants provided consent for publication of study results of the collected biomaterials paired with anonymized information on age, sex, and location. The authors declare that they have no competing interests.

3.8. Supplemental Materials

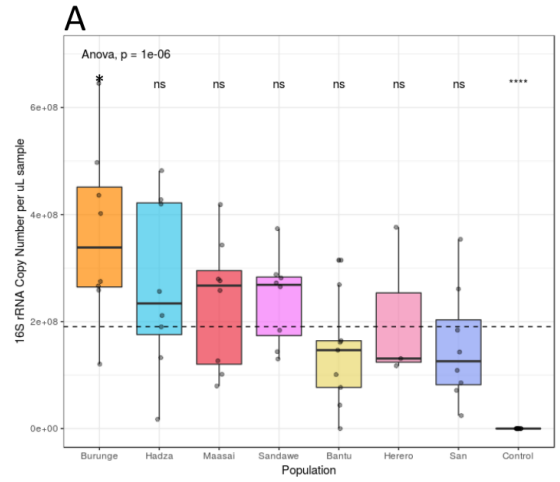
3.8.1. Supplemental figures

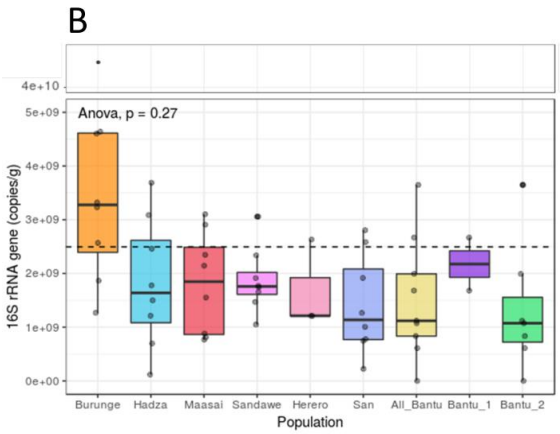


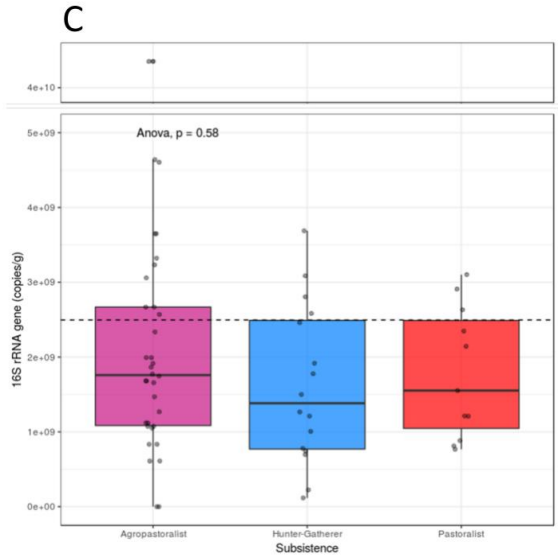
Supp. Figure 3-1. Unassigned reads that were not classified into any known taxa within the kingdom Bacteria. A) The mean unassigned abundance per individual per population. B) The mean number of unassigned OTUs per individual per population. C) The mean number of unassigned reads per individual per population.

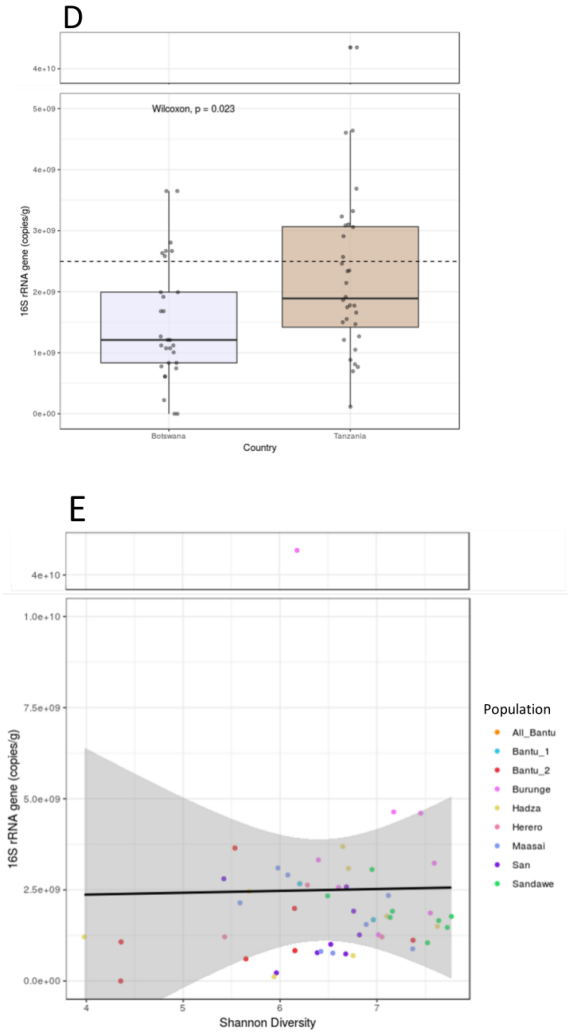


Supp. Figure 3-2. Population Averaged Species Accumulation Curves showing OTUs greater than 0.01% abundance for each population

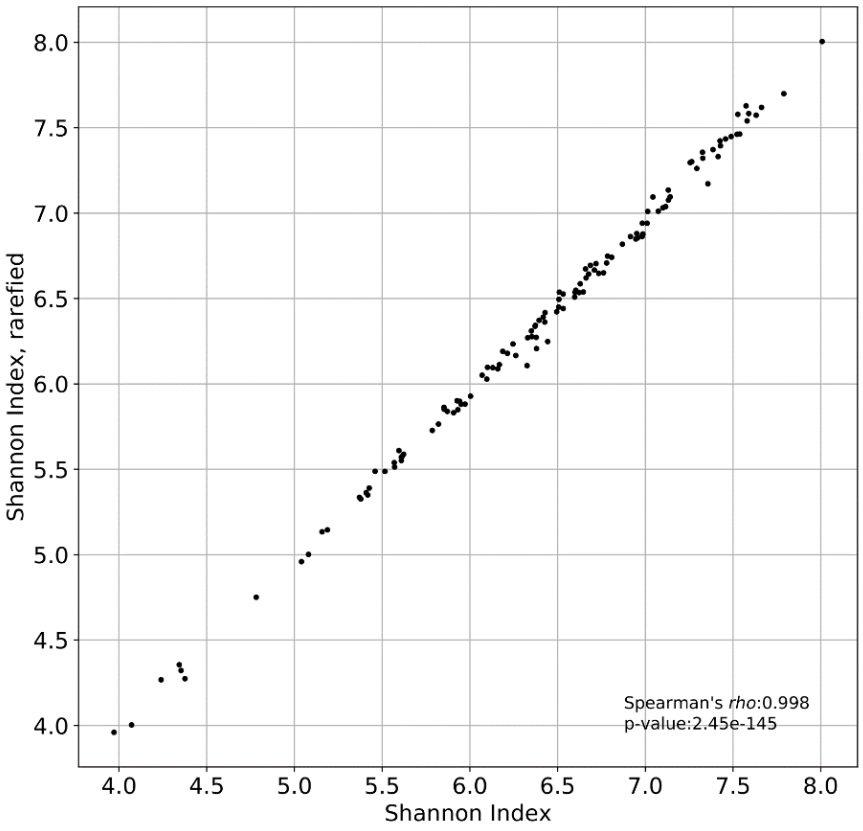

A
Anova, p = 1e-06
*
ns
ns
ns
ns
ns
ns
****
16S rRNA Copy Number per uL sample
6e+08
4e+08
2e+08
0e+00
Burunge
Hadza
Maasai
Sandawe
Bantu
Herero
San
Control
Population


B
4e+10
5e+09
Anova, p = 0.27
16S rRNA gene (copies/g)
4e+09
3e+09
2e+09
1e+09
0e+00
Burunge
Hadza
Maasai
Sandawe
Herero
San
All_Bantu
Bantu_1
Bantu_2
Population

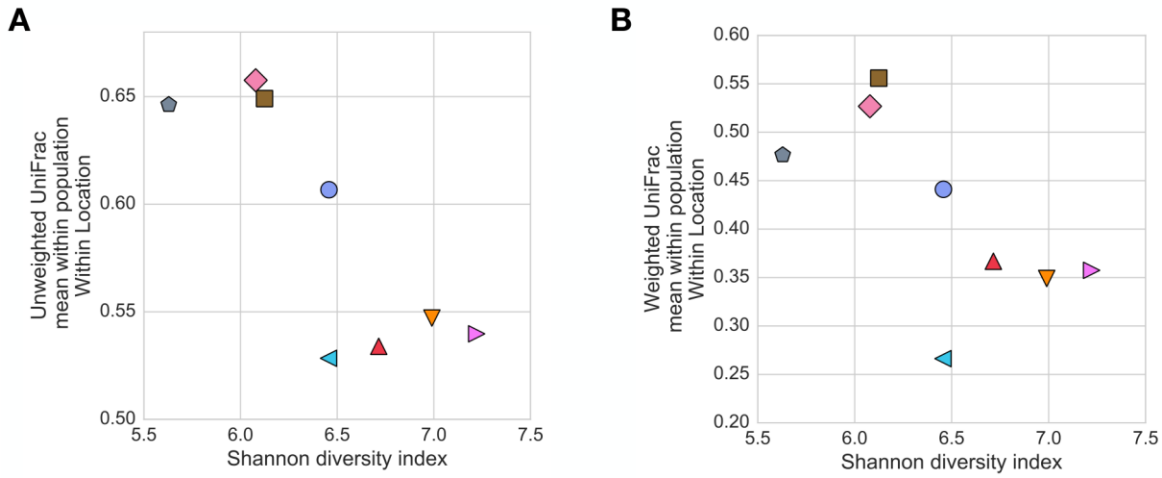

C
4e+10
5e+09
Anova, p = 0.58
16S rRNA gene (copies/g)
4e+09
3e+09
2e+09
1e+09
0e+00
Agropastoralist
Hunter-Gatherer
Pastoralist
Subsistence



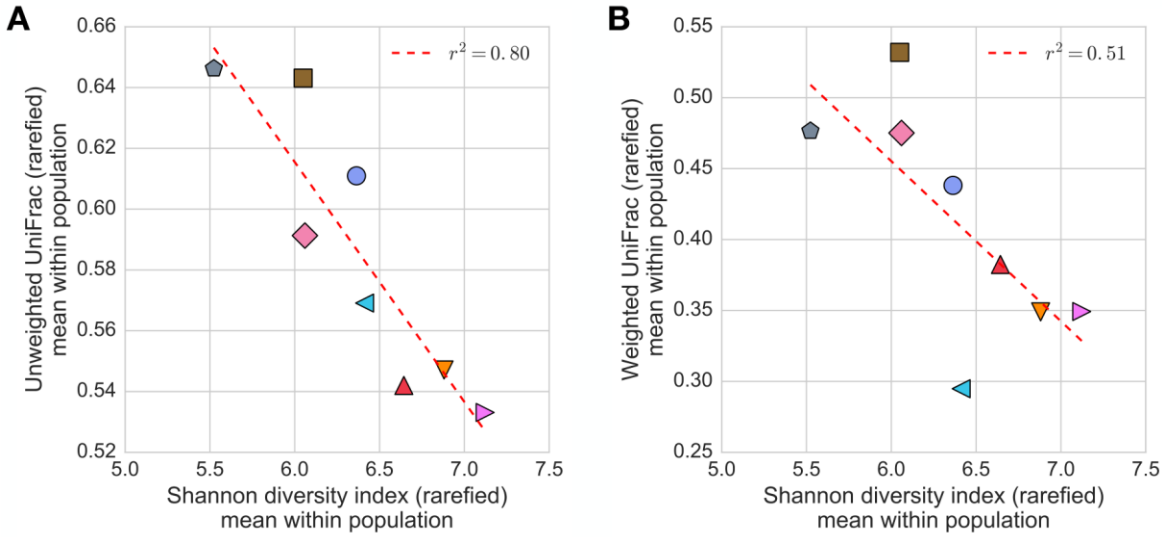
Supp. Figure 3-3. 16S rRNA gene qPCR values across populations, countries, and subsistence groups. A) 16S rRNA gene copy number per uL of sample within each population. Lines through the center of the boxes show the median, and the top and bottom are the 25th and 75th percentiles. Dashed line in center represents the group mean for all populations, against which a global ANOVA for all populations is derived (adjusted p=1e-06). B) 16S rRNA Copy Number per gram of stool within each population. Global ANOVA for all populations is not significant (adjusted p=0.27). Two samples within the Burunge had values high enough that they are plotted separately above the main boxplot (this was also done for all subsequent panels in S15). C) 16S rRNA Copy Number per gram of stool within each African subsistence group. Global ANOVA for all subsistence groups is not significant (adjusted p=0.58). D) 16S rRNA Copy Number per gram of stool within Botswana and Tanzania. Wilcoxon Rank-Sum test indicates that significant differences in copy number exist between the two countries (FDR corrected p=0.023). E) Linear regression on 16S rRNA Copy Number per gram of stool to Shannon Diversity per sample, colored by Population, and showing 95% confidence intervals.



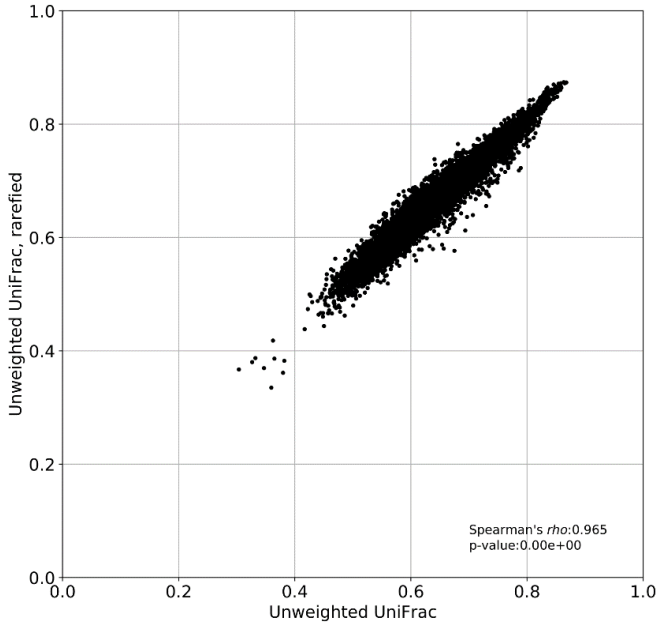
Supp. Figure 3-4. Comparison of the Shannon Index between all reads and rarefied data Read counts subsampled down to 5,000 reads per individual, and reads are shown for all individuals.



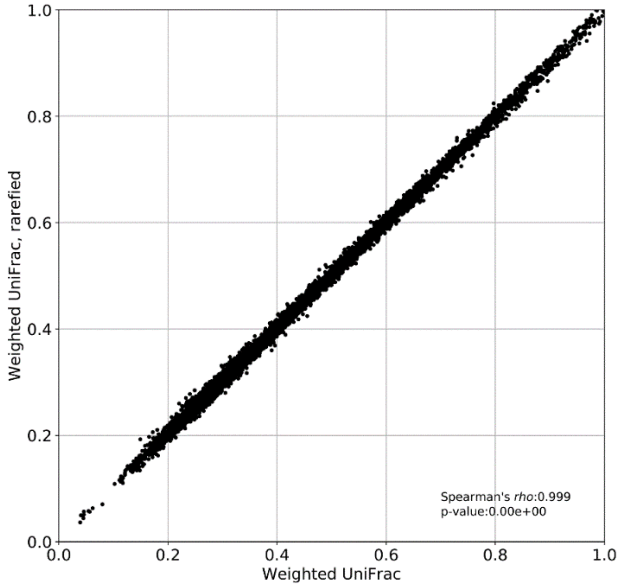
Supp. Figure 3-5. Shannon Index, mean unweighted and weighted UniFrac distances within same sampling location and population. Similar to Figure 2-3 D, E, where the pairs of individuals used for calculating the mean unweighted A) and weighted B) UniFrac distances are restricted to be individuals from the same sampling location as well as from the same population group.



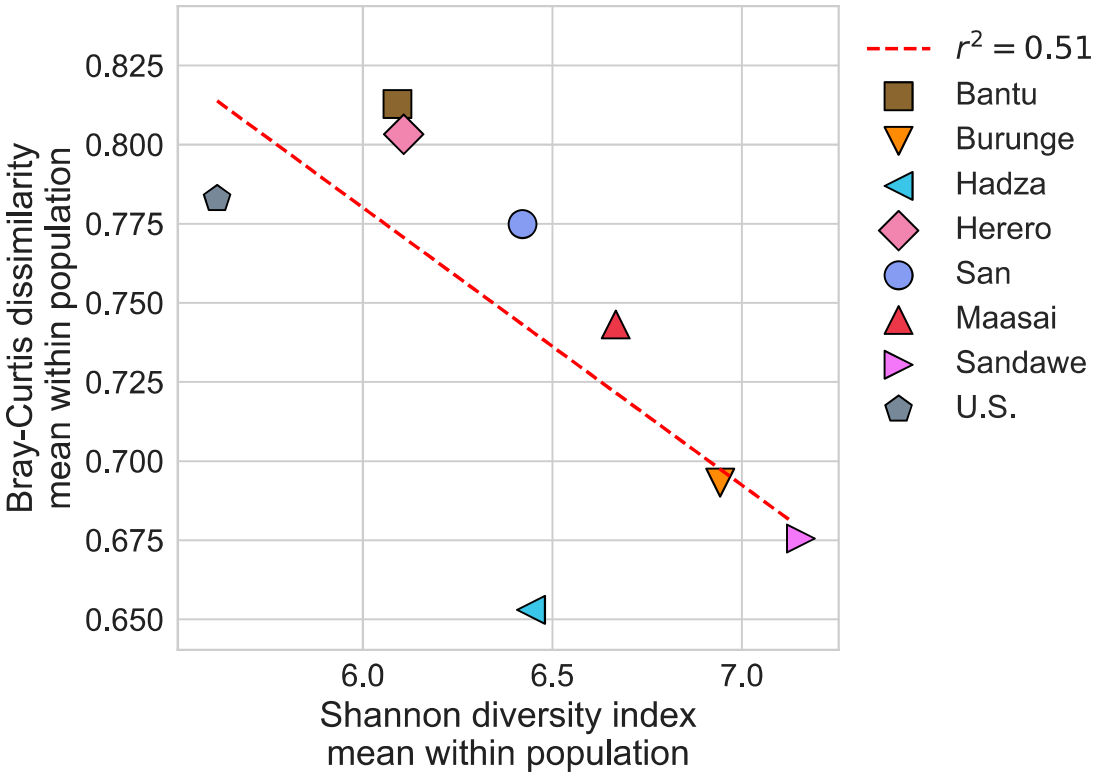
Supp. Figure 3-6. Rarefied Shannon Index, mean unweighted and weighted UniFrac distances within populations. Similar to Figure 2-3 D, E where the OTU data is rarefied to 5,000 counts per individual, showing unweighted A) and weighted B) UniFrac distances.



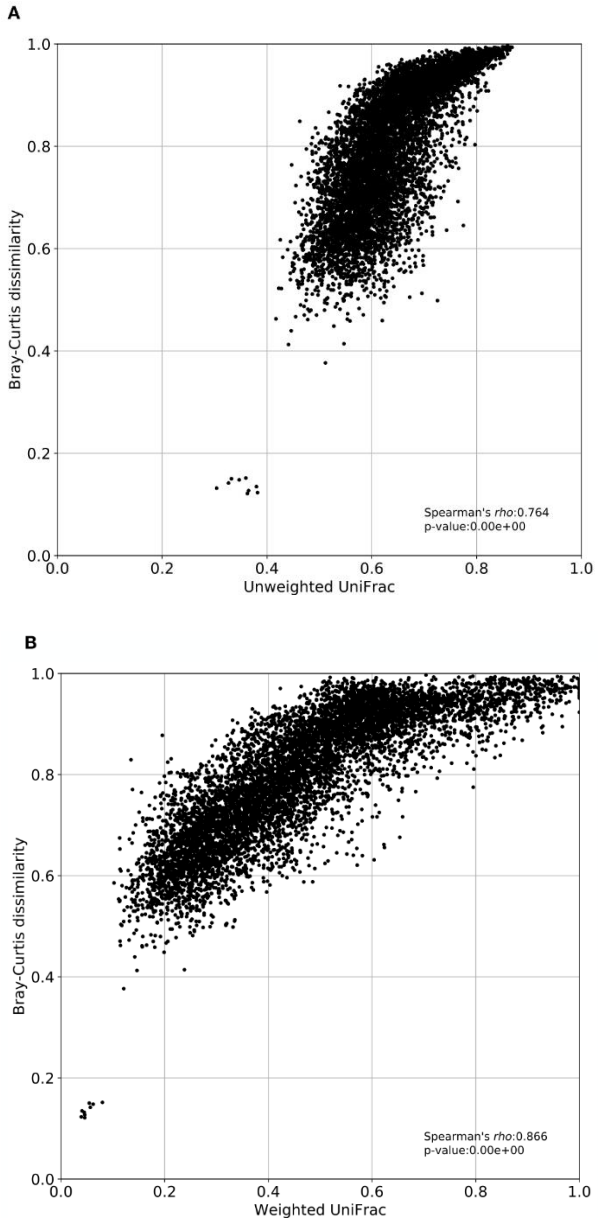
Supp. Figure 3-7. Comparison of the unweighted UniFrac distances between all reads and rarefied data for all pairs of individuals. Read counts subsampled down to 5,000 reads per individual.



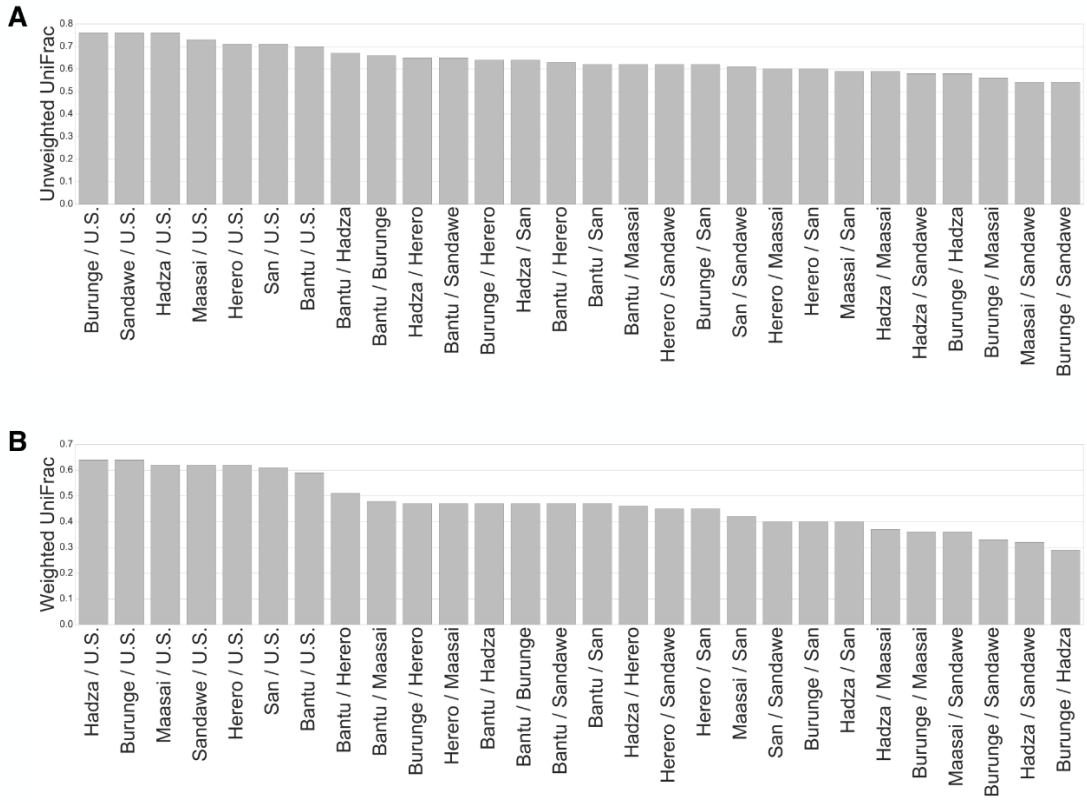
Supp. Figure 3-8. Comparison of the weighted UniFrac distances between all reads and rarefied data for all pairs of individuals. Read counts subsampled down to 5,000 reads per individual.



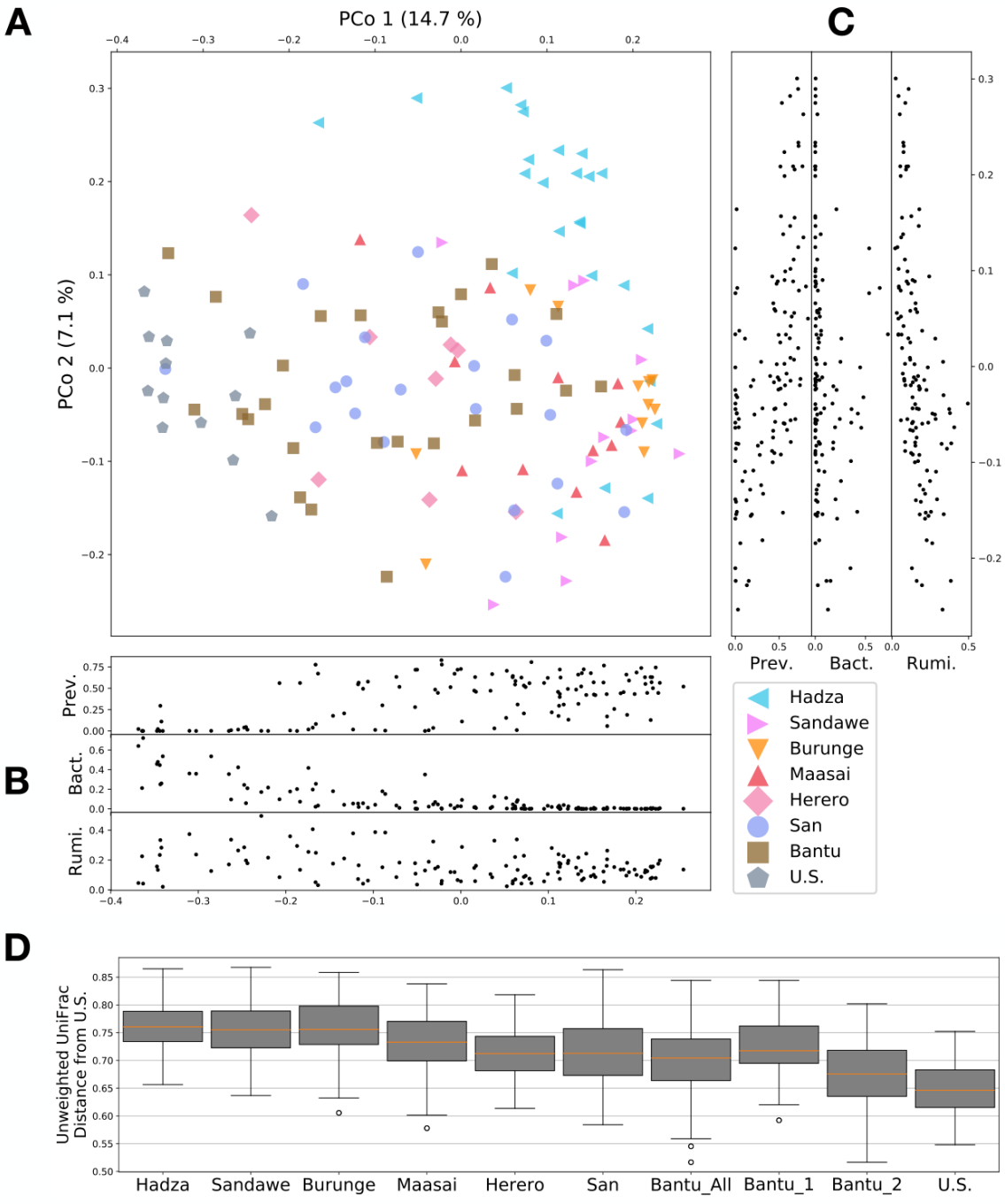
Supp. Figure 3-9: Bray-Curtis dissimilarity measure of beta-diversity within populations Similar to Fig. 2-3 D, E.



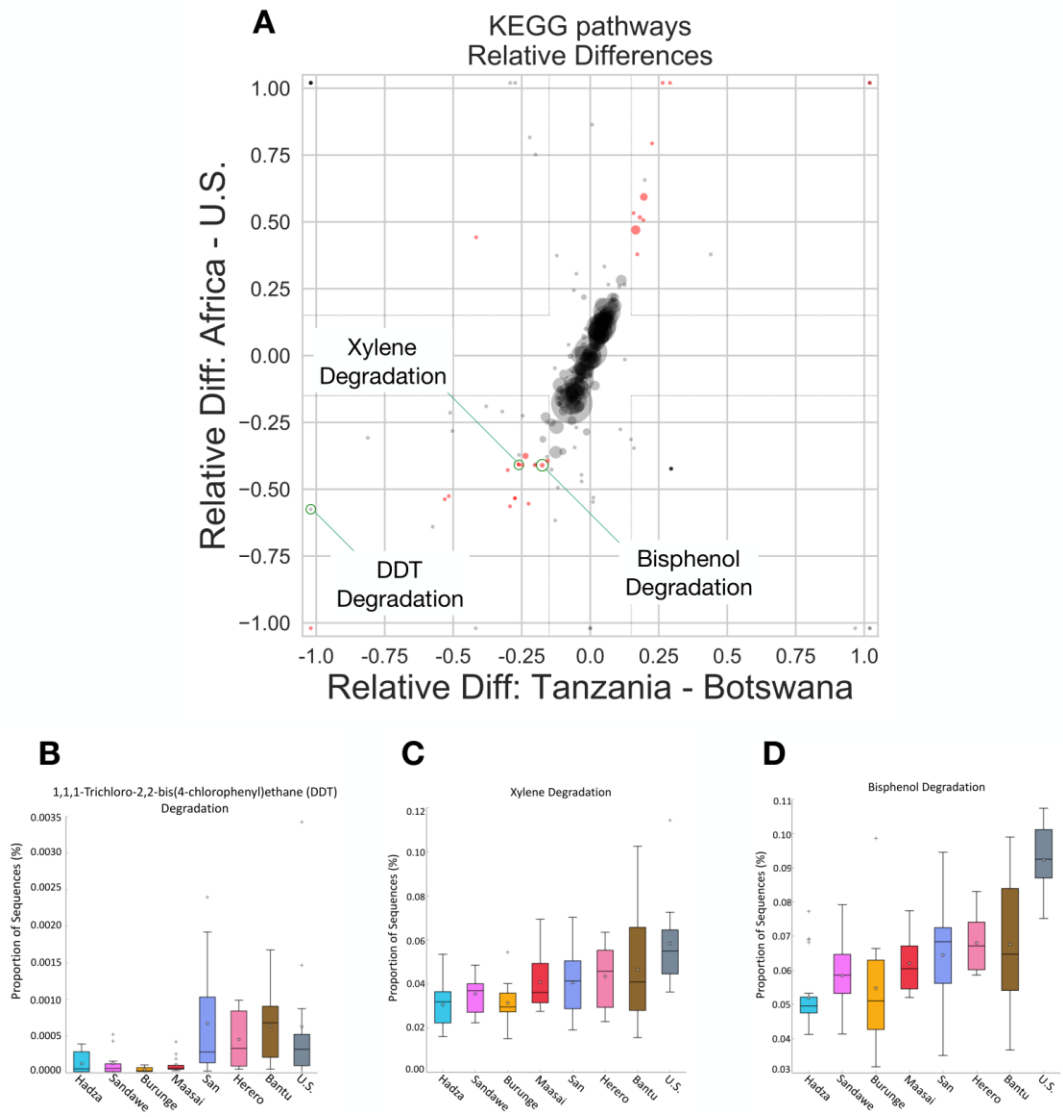
Supp. Figure 3-10. Correlation plot between Bray-Curtis dissimilarity UniFrac distances A) unweighted UniFrac distance and B) weighted UniFrac distance, across all pairs of individuals.



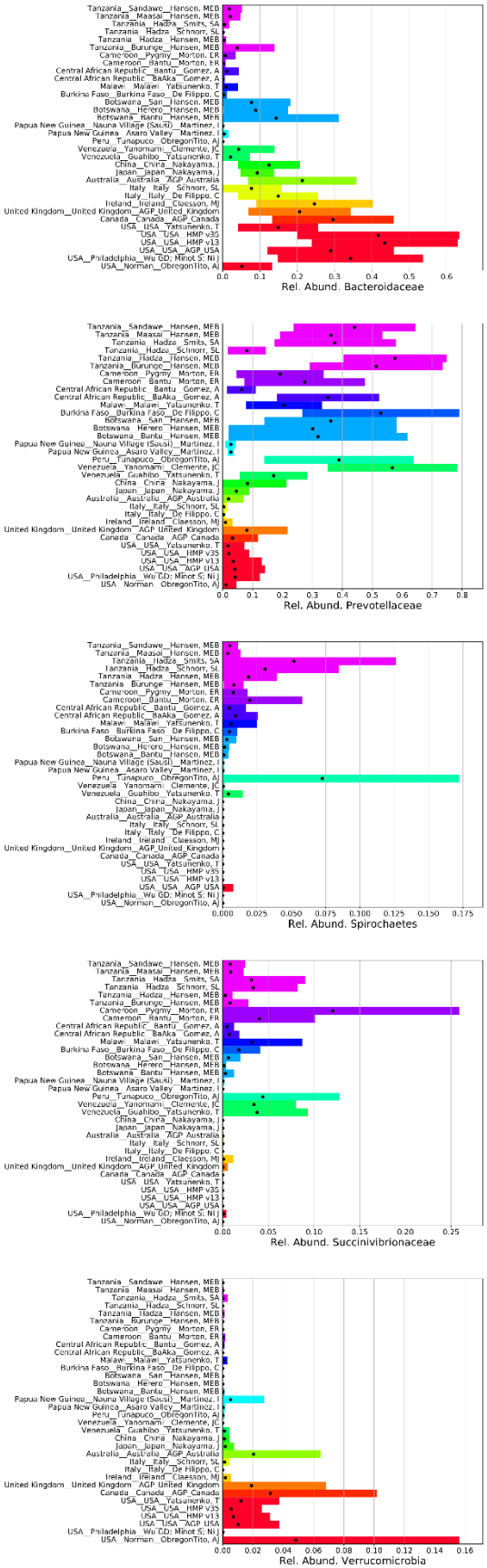
Supp. Figure 3-11. Mean pairwise unweighted (A) and weighted (B) UniFrac distances between populations



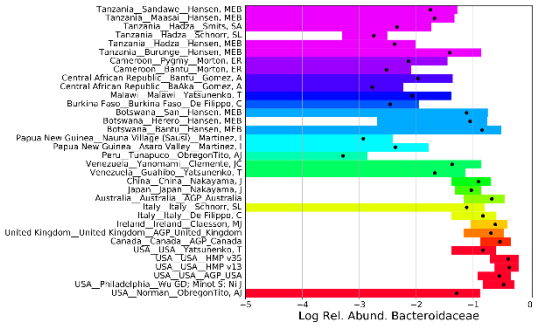
Supp. Figure 3-12. Principal Coordinate Analysis of unweighted UniFrac distances. Analogous to Figure 2-5.

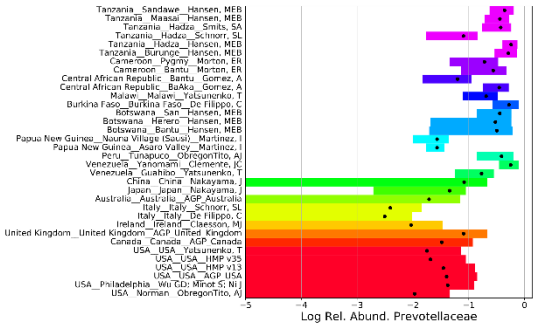


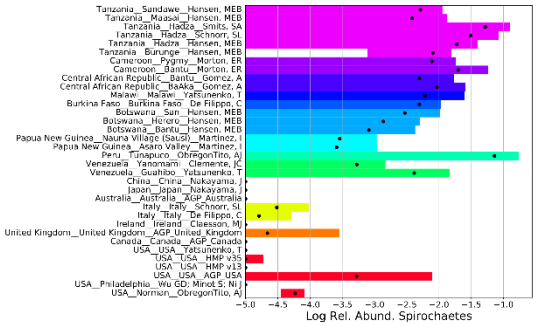
Supp. Figure 3-13: Relative pathway frequency differences between Tanzania and Botswana, and Africa and the U.S., for all KEGG pathways analyzed. Differences between Tanzania and Botswana are shown on the x-axis and differences between Africa and the U.S. are shown on the y-axis for all KEGG pathways analyzed (N=328). Marker size is proportional to the mean pathway frequency across the U.S. and Africa. Dotted lines demark relative differences of ±15%. Pathways in red denote those shown in Table S1 (absolute relative differences > 15% and FDR < 0.1). Panels B), C), and D) show the pathway frequency distribution boxplots per population for the DDT degradation, xylene degradation, and bisphenol degradation pathways, respectively. Lines through the center of each box show the median, the star represents the mean, and the top and bottom are the 25th and 75th percentiles.

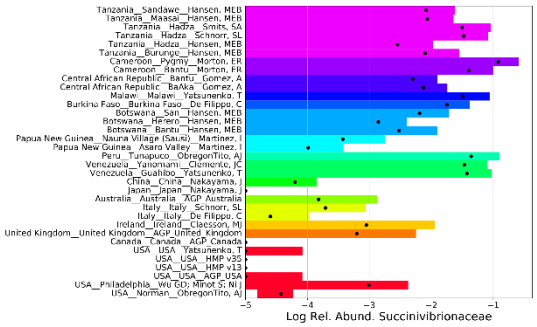


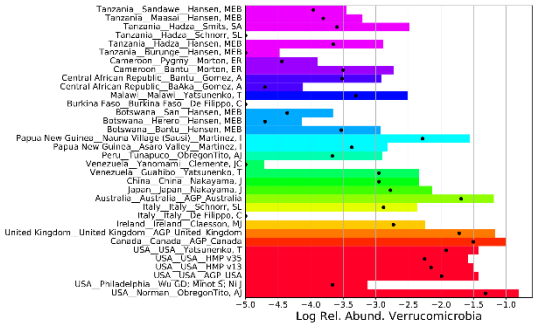
Supp. Figure 3-14. Comparison of reported relative abundances of our study cohort with those of global populations
Means and standard deviations are shown. The taxa included are the bacterial Families *Bacteroidaceae*, *Prevotellaceae*, *Spirochaetaceae* , *Succinivibrionaceae*, and the bacterial Phylum *Verrucomicrobia*, shown on a linear abundance scale and have been adapted from Smits et al. (2017).











Supp. Figure 3-15. Comparison of reported relative abundances of our study cohort with those of global populations on a log scale
Same as Supp. Figure 3-14 but transformed on a log scale.

3.8.2. Supplemental tables and files

Supplemental Tables:

All supplemental tables are attached in the Digital Supplement named

SupplementalTables_Ch3.xlsx.

Supp. Table 3-1. Sample Metadata.

Supp. Table 3-2. Unweighted UniFrac distances between all pairs of populations.

Supp. Table 3-3. Weighted UniFrac distances between all pairs of populations.

Supp. Table 3-4. PERMANOVA results for unweighted UniFrac distances between countries.

Supp. Table 3-5. PERMANOVA results for weighted UniFrac distances between countries.

3-6. PERMANOVA results for unweighted UniFrac distances between all pairs of populations.

Supp. Table 3-7. PERMANOVA results for weighted UniFrac distances between all pairs of populations.

Supp. Table 3-8. PERMANOVA results for unweighted UniFrac distances between groups defined by subsistence category.

Supp. Table 3-9. PERMANOVA results for weighted UniFrac distances between groups defined by subsistence category.

Supp. Table 3-10. PERMANOVA results for unweighted UniFrac distances between sexes within the Maasai, Hadza, San, and U.S. populations.

Supp. Table 3-11. PERMANOVA results for weighted UniFrac distances between sexes within the Maasai, Hadza, San, and U.S. populations.

Supp. Table 3-12. PICRUSt analysis of pathway abundance differences between Tanzania and Botswana, and the U.S. and Africa.

Supp. Table 3-13. PICRUSt analysis of pathway abundance differences between the Hadza population and the remaining Africans.

Supp. Table 3-14. PICRUSt analysis of pathway abundance differences between the Burunge population and the remaining Africans.

Supp. Table 3-15. PICRUSt analysis of pathway abundance differences between the San population and the remaining Africans.

Supp. Table 3-16. PICRUSt analysis of pathway abundance differences between the Bantu population and the remaining Africans.

Supp. Table 3-17. BMI correlation with Shannon Index, shown for different subsets of individuals included in the correlation.

Supp. Table 3-18. BMI correlation with OTU abundances.

Supp. Table 3-19. BMI correlation with taxa (genus) abundances.

Supp. Table 3-20. Spearman's Rho correlation between host Identity-by-Descent and Bacterial UniFrac distance, among all pairs of individuals of the same sex within each population.

Supp. Table 3-21. Linear regression analysis of the dependence of UniFrac distance on host relatedness and host geographic separation.

Supp. Table 3-22. Modified gut microbiome meta-analysis from Smits et al. (2017).

Supp. Table 3-23. Dietary literature review on sub-Saharan African populations included in this study.

Supp. Table 3-24. Closest towns to Tanzanian and Botswanan sampling sites showing estimated sample sizes from census data for both sampling sites and closest town, location inhabitant groups, and corresponding inhabitant number created from the

United Nations Statistics Division, and distance in kilometers of the town from the sampling site.

Supp. Table 3-25. 16S rRNA gene sequencing metadata and amplification conditions.

Supp. Table 3-26. Sequences of denovo OTUs specifically named in the paper and supplemental figures.

Supp. Table 3-27. Individual counts for all 17,870 OTUs with a mapped bacterial lineage used.

Supp. Table 3-28. PICRUSt quality control metric Nearest Sequenced Taxon Index (NSTI) for each individual.

Supp. Table 3-29. Two group KEGG pathway abundance differences.

Supplemental Files:

All supplemental files are attached in the Digital Supplement named

SupplementalFile1_Ch3.ibd, SupplementalFile2_Ch3.id, and SupplementalFile3_Ch3.gz.

Supplemental File 1*: *The estimated identity-by-descent used for the within-population correlation between UniFrac distance and host relatedness, as calculated using the plink --genome routine on the Illumina Omni 5M genotype array dataset for 97 African individuals.

Supplemental Files 2 and 3: Genetic relationship matrix file sets used for linear regression analysis of UniFrac dependence on host relatedness and host geographic separation.

This is constructed using the GCTA software (--make-grm), based on Illumina Omni 5M genotype array dataset for 97 African individuals. Supplemental File 2 contains the (anonymized) individual identification codes. Supplemental File 3 contains the genetic similarity values, where each row corresponds to a pair of individuals.

Chapter 4. Gut microbiome composition is predictive of gastrointestinal parasites in Cameroonians

Rubel, M.A., Abbas, A.A., Taylor, L.J., Mbunwe, E., Connell, A. Tanes, C., Bittinger, K., Fokunang, C., Njamnshi, A.K., Bushman, F.D., Tishkoff, S.A.

4.1. Abstract

African populations provide a unique opportunity to interrogate host-microbe co-evolution and its impact on adaptive phenotypes, thanks in part to their genomic, phenotypic, and cultural diversity. Here, we integrate sequence data from gut amplicon and metagenomic sequencing with data on parasite burden, and immune parameters for 575 ethnically diverse Africans from Cameroon who follow pastoralist, agropastoralist, and hunter-gatherer lifestyles. These populations were compared to an urban U.S. population from Philadelphia. We found that Fulani pastoralists who consume high levels of milk possessed an enrichment of bacteria that catabolize galactose, an end product of lactose metabolism. We assessed these data for interactions between eukaryotic protists and prokaryotes (transkingdom interactions). Microbiota composition and levels of soil-transmitted helminths varied significantly among Cameroonian populations. Hunter-gatherers had high frequencies of parasite infections, while agropastoralists and pastoralists showed lower levels of parasite infections. Across all populations, increased frequency of gut parasites correlated with increased gut microbial diversity. *Ascaris lumbricoides, Necator americanus, Trichuris trichiura*, and *Strongyloides stercoralis* soil-transmitted helminths (STH), referred to as the "ANTS" group, significantly co-occurred. Gut microbiome composition predicted ANTS positivity with 80% accuracy. Infection with ANTS parasites, in turn, was associated with elevated levels of TH1, TH2, and proinflammatory cytokines, indicating engagement of multiple immune mechanisms. These data document transkingdom interactions that are correlated with distinctive host immune responses and indicate that Cameroon pastoralist populations may possess unique bacteria to facilitate digestion of dairy products.

4.2. Introduction

Twenty-four percent of the world's population, predominantly in developing countries, is estimated to be infected with gastrointestinal parasites. Enteric parasites are understudied components of the complex ecosystem of microorganisms that can colonize the human gastrointestinal tract, and their effects on host physiology and the gut microbiota remain poorly understood. The mechanisms by which heterogeneity in microbial communities

modulates immune responses to infectious diseases is just beginning to be studied. Industrial countries are characterized by an overall reduction in exposure to pathogens and microbes consequent to cultural and technological societal shifts. Microbial diversity may have an essential role in host physiology. Thus, there is a need to characterize microbial diversity from rural populations living traditional lifestyles before they make subsistence shifts to industrialization and this diversity is lost (Bello et al., 2018).

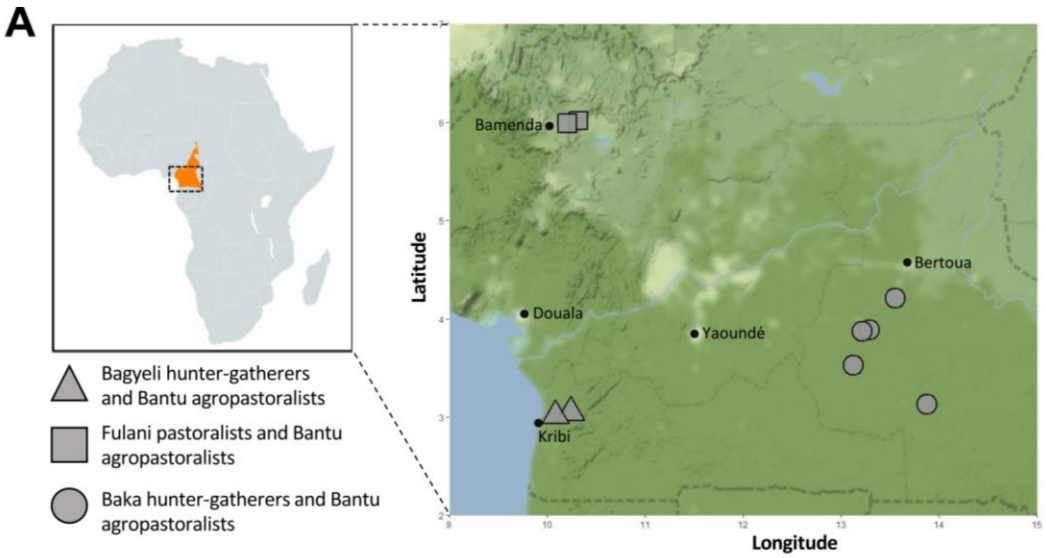
The relative contribution of environmental and physiological variables to development of gut microbial composition, including geography, pathology, diet, immune parameters, genetics, and ethnicity are incompletely understood. The nature of a healthy gut microbiome differs among ethnic groups, locations and lifestyles (Ayeni et al., 2018; Gomez et al., 2016; Hansen et al., 2019; Schnorr, 2018; Yatsunenko et al., 2012). For example, different bacterial genera tend to dominate the fecal microbiome in people around the world living traditional rural lifestyles (*Prevotella*) compared to urban dwellers (*Bacteroides*) (Arumugam et al., 2011; Gomez et al., 2016; Hansen et al., 2019; Yatsunenko et al., 2012). Recent research has shown that infection with gastrointestinal parasites *Entamoeba histolytica* (Morton et al., 2015), *Ascaris lumbricoides* (Rosa et al., 2018), *Necator americanus* (Jenkins et al., 2017; Rosa et al., 2018), and *Trichuris trichiura* (Cooper et al., 2013; Lee et al., 2014; Rosa et al., 2018) influence the structure and function of the human gut microbiome. Microbiota-parasite interactions within the human gut may have profound impacts on the course of systemic infection, parasite virulence, and host immune response (Leung, Graham, & Knowles, 2018). Prior microbiome research has considered many gastroenteric parasites as separate exposures

(Lee et al., 2014; Lin et al., 2013; Morton et al., 2015; Rosa et al., 2018). However, in developing countries, parasite coinfection is frequent (Morton et al., 2015), and likely has a role in microbiome structure and function. Here we focus on populations from Cameroon that are genetically, linguistically, phenotypically, and culturally diverse. These populations have different types of subsistence practices but share overlapping environments and high infectious disease burdens. Integrating metagenomic comparisons of the microbiota, quantitative measurements of multiple parasite loads, host immune parameters, and extensive demographic data, we investigate correlations between microbial structure and function, host immune response, and parasite infections.

4.3. Results

4.3.1. Data collection and populations studied

The Cameroonian populations studied here are Mbororo Fulani pastoralists (hereafter referred to as "Fulani"), Baka and Bagyeli rainforest hunter-gatherers, and Bantu-speaking agropastoralists (hereafter referred to as Bantu). Fulani pastoralists have subsistence practices centered largely around cattle, and the Bantu grow crops and raise livestock. The Baka and Bagyeli hunter-gatherers, who are sometimes referred to as "pygmies" for their short stature, engage in small-scale agriculture but also forage for meat and plant materials. These populations were sampled over nine sites in the Northwest, South, and East regions of Cameroon (Figure 4-1).


A
Bamenda
Bertoua
Douala
Yaoundé
Kribi
Latitude
Longitude
Bagyeli hunter-gatherers and Bantu agropastoralists
Fulani pastoralists and Bantu agropastoralists
Baka hunter-gatherers and Bantu agropastoralists


B

C

Country	Population	Subsistence	Number Total	Number Female	Number Male	Number Sex Unknown	Age Ave.	Age Min.	Age Max.
Cameroon	Bagyeli	Hunting & Gathering	33	20	13	0	33.7	6	67
	Baka	Hunting & Gathering	117	67	50	0	31.9	5	76
	Bantu	Agropastoralist	322	197	125	1	43.8	4	90
	Mbororo Fulani	Pastoralist	103	56	47	0	38.6	7	87
		subtotals:	575	351	240	1	36.9	4	90
U.S.	Philadelphia	Industrial agropastoralist	37	25	12	0	34.2	22	78
		Totals:	**612**	**376**	**235**	**1**	**35.55**	**4**	**90**

Figure 4-1. Sampling sites and demographics (Cameroon and U.S.)
A) Cameroon sampling sites. At every sampling site, Bantu agropastoralists were sampled. One of two hunting and gathering groups (Baka and Bagyeli) or Fulani pastoralists were also sampled at these sites. B) Top- Image of a representative village with a large population of Bantu agropastoralists (creative commons license). Bottom- Image of a representative village with a large population of hunter-gatherers (photograph by Meagan A. Rubel). D) Demography table of truncated metadata for Cameroon and U.S. (See Supp. Table 1 for full metadata).

Because subsistence and ethnicity are strongly correlated in these populations (e.g., all pastoralists are Fulani, all agropastoralists are Bantu, and all Baka and Bagyeli are hunter-gatherers), only subsistence was included as a statistical parameter in the analyses described below. Extensive anthropometric measurements and biomarkers (Supp. Table 4-1) were recorded for Cameroonians (see Methods). We also included two cohorts of U.S. industrial agropastoralists who have diets high in animal fats, proteins, and refined, processed foods that are the byproducts of intensive agricultural and pastoral practices, from the Human Microbiome Project (Consortium, 2012; Human Microbiome Project Consortium, 2012) and the COMBO study (Wu et al., 2011). By integrating multiple data

types, we define complex interactions between parasites, the microbiota, and their human hosts across populations and lifestyles.

4.3.2. Quantification of pathogens and their correlates with host physiology

Using thin and thick smears with light microscopy, a total of 198 blood infections were identified for *Plasmodium falciparum*, *Microfilaria loa loa*, *Mansonella perstans*, *Wuchereria bancroftii*, and *Microfilaria* spp.; *P. falciparum* accounted for the most infections detected by microscopy (N=96) (Supp. Figure 4-1).

Fecal parasites were identified by light microscopy using wet-mount techniques and qPCR (Mejia et al., 2013). The concordance between microscopy and qPCR results for fecal parasite infection was 86% (Supp. Table 4-1). qPCR was more sensitive, detecting almost three times more infections than microscopy. We tested for the following fecal parasites by qPCR: the giant roundworm *Ascaris lumbricoides*, hookworms *Necator americanus* and *Ancylostoma duodenale*; whipworm *Trichuris trichiura*; roundworm *Strongyloides stercoralis*; and protists *Giardia lamblia*, *Entamoeba histolytica*, and *pan-Cryptosporidium* spp. (Supp. Figure 4-2). None of our samples were positive for *A. duodenale*, and it was not considered in further analyses. Given the increased sensitivity, qPCR-confirmed parasite infections were used for subsequent analysis.

Total infectious disease burden was determined by counts of fecal/blood parasites. This analysis indicated that Cameroonian populations had significantly different distributions of infectious disease burden (Kruskal-Wallis chi-squared = 244, df = 3, p-value < 2.2e-16). All groups were significantly different from each other in terms of their distribution

for pathogens except for the Baka and Bagyeli who were statistically indistinguishable from each other (FDR-corrected p-values<0.05, Dunn's test of multiple comparisons). The Bagyeli hunter-gatherers had the highest individual infection rate (an average of 3.91 infections/person) compared to other Cameroonian populations (Baka = 2.83, Fulani = 0.22, Bantu = 1.13) (Fig. 2A).

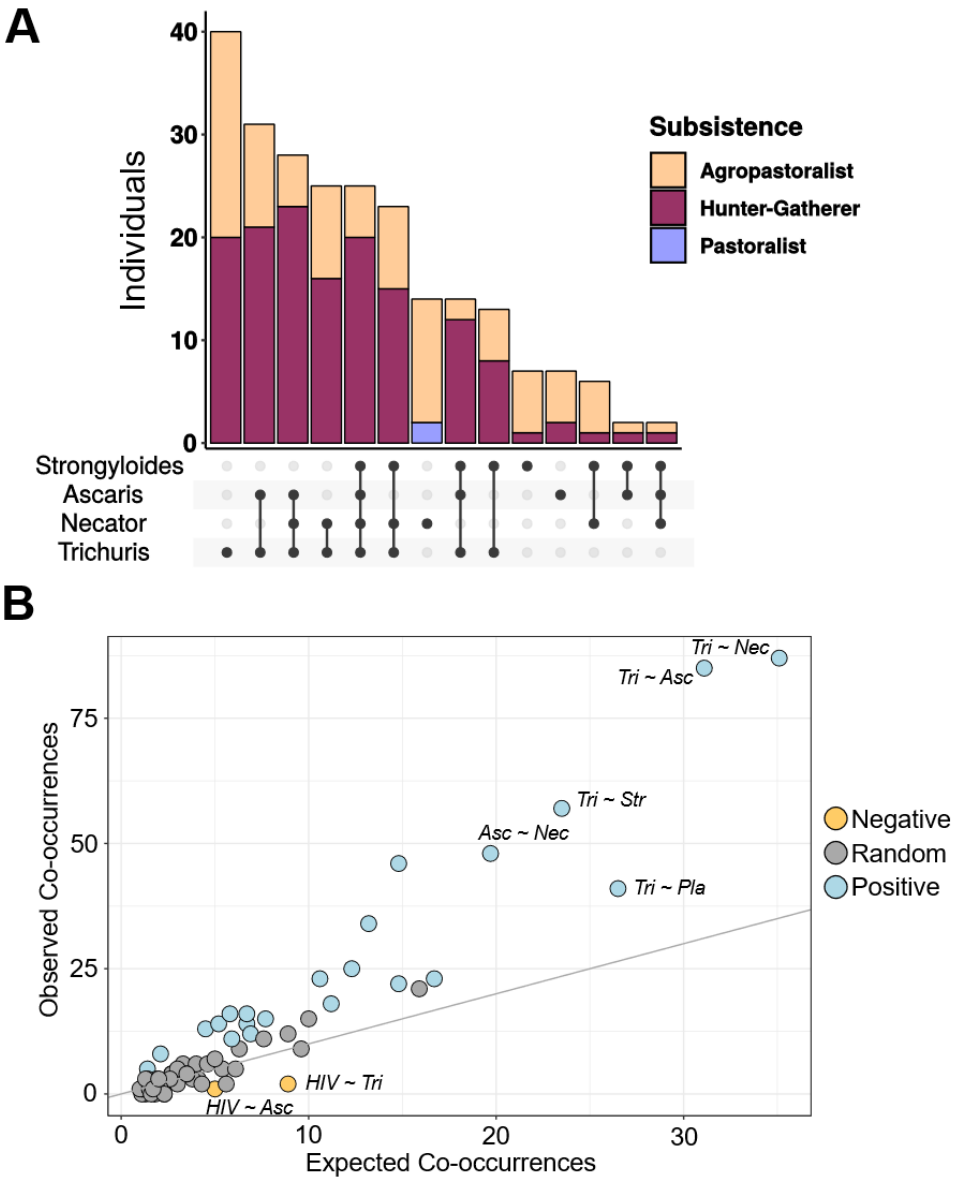


Figure 4-2. Pathogen characterization in the Cameroon cohort
A) Distribution of individuals infected with different combinations of soil-transmitted helminths within the "ANTS" group in the Cameroon sample colored by subsistence group. B) Positive, negative, and random association based on a probabilistic model of pathogen co-occurrence, calculated across all types of tested pathogens (blood parasites, fecal parasites, and HIV) for all Cameroonians. The expected frequency is calculated from presence/absence data of parasites, with the assumption that each distribution of each pathogen is independent and random of other pathogens. Probabilities that are more extreme than would have been obtained by chance are shown for positive co-occurrences in blue (as one pathogen occurrence increases, the other increases), negative co-occurrences in yellow (as one pathogen occurrence increases, the other decreases) and random co-occurrences in grey (no significant association). Top hits for negative and positive co-occurring pathogens are annotated as follows: Asc: *Ascaris lumbricoides*, HIV*:* Human Immunodeficiency Virus, Nec*: Necator americanus*, Pla*: Plasmodium falciparum*, Tri*: Trichuris trichiura*.

Concurrent infection with >1 tested pathogen occurred in (226/575) 39% of Cameroonians. We performed species co-occurrence analysis to identify combinations of HIV, blood, and fecal pathogens that occurred together more frequently than expected by chance. Twenty-one pathogen pairs positively co-occurred more than expected by chance (p-values < 0.05 by the hypergeometric distribution). The parasites that most frequently co-occurred were *A. lumbricoides, N. americanus, T. trichiura*, and *S. stercoralis* (the occurrence of any these parasites is labeled as "ANTS" group) (p-values < 0.05 by hypergeometric distribution) (Figure 4-2B, Supp. Table 4-2).

To determine which physiological variables were intercorrelated, we generated correlation coefficients and tested their significance (Wei & Simko, 2017). Co-occurrence results indicated that eosinophils, white blood cell count (WBC), and temperature were significantly positively correlated with any metadata variable that incorporated ANTS helminth parasites (Spearman's correlation test, corrected p-values < 0.01, Supp. Table 4-3, Supp. Figure 4-3). Other significantly positive correlates included

WBC with eosinophils, and temperature with WBC (Supp. Table 4-3, Supp. Figure 4-3). Eosinophilic leukocytes are normally a small fraction of WBC but increase dramatically during helminth infection (Huang & Appleton, 2016). The other WBC fractions that were tested in correlation analysis (lymphocytes, monocytes, neutrophils), which represent most of the remaining portion of total WBC, were significantly negatively correlated with all variables incorporating ANTS (p-values < 0.01). Furthermore, they were significantly negatively correlated with all blood parasites tested in this study. Since many individuals are concurrently infected with ANTS and blood parasites, the positive associations of eosinophils and any variable including ANTS may outweigh other WBC cell responses to blood parasite (or other) infections.

A combination HIV-1/2 immunoassay was conducted on the plasma of 524 individuals who had produced a fecal sample, which showed that 28 participants were HIV positive. No parasites were positively associated with HIV status by this assay. HIV co-occurred less frequently than expected by chance with *T. trichiura* and *A. lumbricoides* (Figure 4-2B); however, this observation is confounded by the distribution of the parasites and HIV+ status by ethnic group (Supp. Table 4-1), and is unlikely to be a direct consequence of HIV infection.

4.3.3. Association of pathogens with the gut microbiome

We amplified and sequenced the V4 region of the bacterial 16S rRNA marker gene from fecal samples obtained from 103 Fulani, 150 Baka and Bagyeli, 322 Bantu, and 37 U.S. industrial agropastoralists (COMBO cohort). A comparison of alpha-diversity metrics

(Faith's phylogenetic diversity and bacterial richness) indicated that that the Bagyeli had the highest diversity and richness, and that the Fulani agropastoralists had the lowest diversity and richness, with the U.S. being the second lowest (Supp. Figure 4-4A-D). In the Cameroonian samples, we observed a significant positive correlation between increasing numbers of parasite infections and higher phylogenetic alpha diversity (Figure 4-3A) (One-way ANOVA, p-values < 0.01).

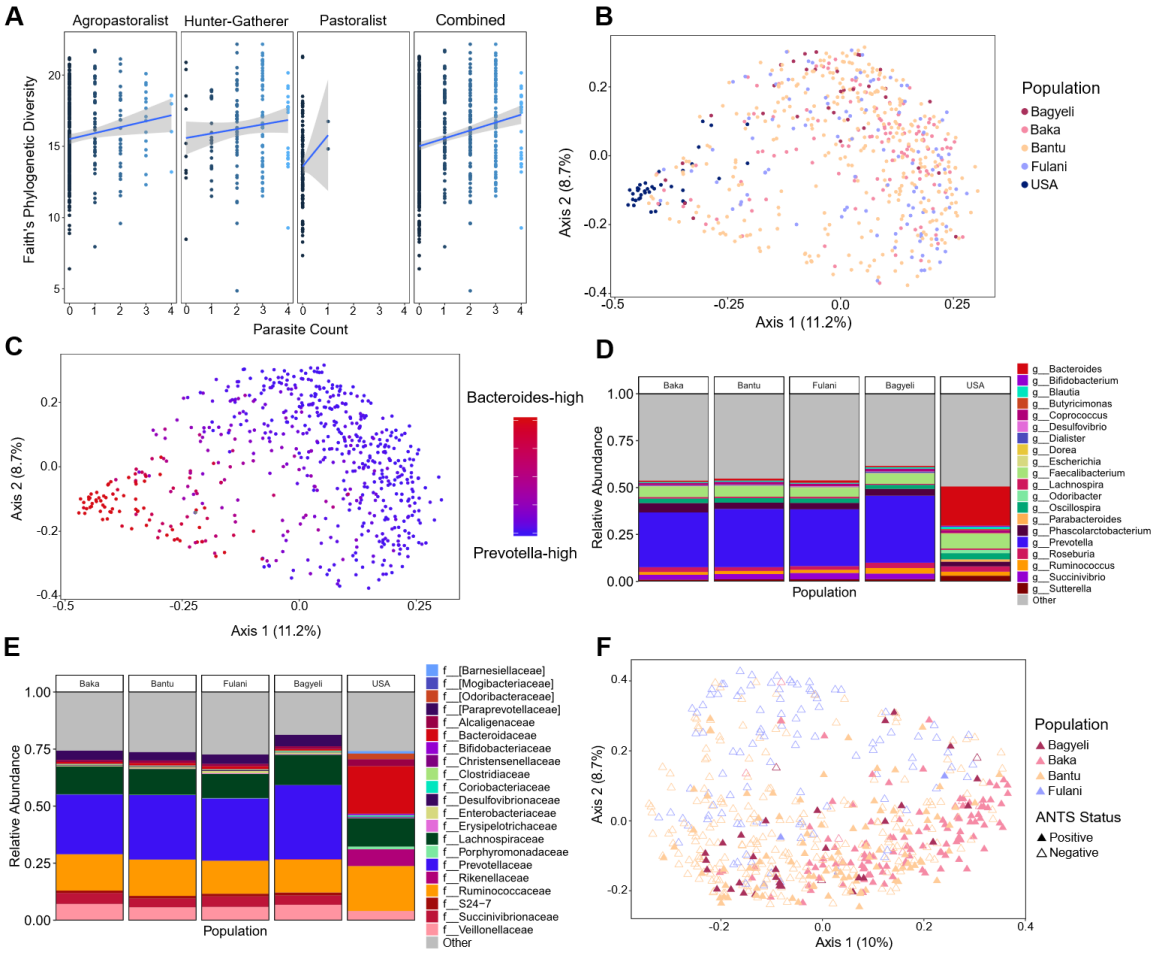


Figure 4-3.16S rRNA Diversity and Composition of the Gut Microbiome and Parasites A) Total bacterial diversity measured against total "ANTS" parasite count, where parasite count is a presence/absence count of unique instances of a parasite. Thus, an individual with a score of "4" has qPCR-confirmed infection for four different gastrointestinal parasites (see Supp. Table 1 for details). Bacterial diversity increases with gastrointestinal parasitemia for all Cameroonian subsistence groups. B) Bray-Curtis distances on 16s rRNA V4 bacterial abundances show that most Cameroonians cluster separately from U.S. individuals (PERMANOVA p = 0.001). C) 16s rRNA V4 clustering of U.S. versus most Cameroonians along separate axes is largely reflective of differences between two highly abundant genera, *Prevotella* and *Bacteroides*. Several pastoralists and agropastoralists overlap with the U.S. samples. Age and sex clusters were not significant by PERMANOVA. D) Fifteen most abundant bacterial genera per population, averaged across populations, studied in 16S analysis. E) Fifteen most abundant bacterial families per population, averaged across populations, studied in 16S analysis. F) Bray-Curtis on Cameroonians, showing 16s rRNA V4 bacterial abundances colored by "ANTS" parasite positivity. ANTS positive samples are significantly different from ANTS negative samples by PERMANOVA (p = 0.001) across all Cameroonians. ANTS positive samples remain significantly different by PERMANOVA from ANTS negative samples when only considering Bantu agropastoralists (Supp. Fig. 10B), who are the only individual population with large cohorts of both positive and negative individuals in this study (p = 0.001).

The agropastoralists were the only subsistence group that had a significant positive correlation between bacterial diversity and ANTS count in a linear regression model (*t*-test with FDR corrected p-value = 0.02). Any single ANTS parasite sufficed to confer increased diversity (One-way ANOVA with Tukey's post-hoc testing; adjusted p-values < 0.05). Across the entire Cameroonian cohort, bacterial diversity significantly increased by 0.55 for every additional ANTS helminth (*t*-test, p =1.3x10-7). Increased bacterial diversity with a single helminth or protist parasite infection has been noted before in populations from Cameroon (Morton et al., 2015), Indonesia (Lin et al., 2013; Rosa et al., 2018), Liberia (Rosa et al., 2018), Bangladesh (Lin et al., 2013), and Malaysia (Lee et al., 2014). However, this is the first time that significantly co-occurring gastroenteric parasites have been shown to additively associate with increased bacterial diversity.

Bacterial microbiome composition in most individuals from Cameroon (based on both abundance and presence/absence) was significantly different from the microbiome composition of U.S. individuals (Figure 4-3B) (PERMANOVA p-value = 0.003 for both) (Fig. 3B) (Supp. Tables 4-4, 4-5). This observation is largely explained by differential prevalence of *Prevotella* and *Bacteroides* in the U.S. versus Cameroonian populations. *Prevotella* was the most highly abundant bacteria in most Cameroonian individuals, whereas *Bacteroides* was most abundant in the U.S. individuals (Figure 4-3C, 4-3D, 4-3E). An analysis of the top ten significant metadata variables by PERMANOVA (FDR corrected p-values < 0.05, Supp. Tables 4-4, 4-5) revealed that parasite variables and subsistence categories (including total ANTS and parasite counts, being positive for *Trichuris* and *Ascaris,* and Subsistence) explained the most variance in gut composition (largest R^2 values).

ANTS positive individuals also had significantly different bacterial composition compared to ANTS negative individuals when considering all Cameroonians (Figure 4-3F) (PERMANOVA test, p-values < 0.006; Supp. Tables 4-4, 4-5). Given the uneven distribution of ANTS helminth infection among the different groups, we repeated the analysis considering only the Bantu agropastoralists, who had sufficient ANTS cases and controls for statistical comparison and found a similar result (PERMANOVA test, p-values < 0.006). A possible confounding factor is that ANTS infection is more common in Bantu populations from the South and East than the Northwest, so we looked for differences in the gut microbiomes between Bantu individuals by region. Bacterial abundances were still significantly different based on ANTS infection (p = 0.004) in a

PERMANOVA model accounting for both local geography and parasite infection. The majority of differences among Bantu populations were driven by taxa in the *Firmicutes* and *Bacteroidetes* phlya (Supp. Figure 4-5).

We used a supervised machine learning technique, random forest classifiers (RFCs), to determine which gut microbiota best predict metadata variables identified as significantly associated with bacterial prevalence or abundance by PERMANOVA in the pooled Cameroonian dataset. Previous studies indicate that moderate to heavy parasite load is associated with increased morbidity (Hotez et al., 2008) and could affect gut microbiome composition Therefore, we binned individuals based on the highest quartile of qPCR copy number for any of the four ANTS parasites (referred to as "highly positive for ANTS"). RFC analysis revealed that the country of origin could be predicted from microbiome composition with ~90% accuracy, followed by individuals who were highly positive for ANTS (83.97% accuracy). Pastoralist and hunter-gatherer subsistence, as well as infections with (separately) *Microfilaria* spp., *A. lumbricoides*, or *T. trichuris* could be predicted with 81-82% accuracy. Finally, positivity for any ANTS parasite could be predicted with ~81% accuracy (Supp. Table 4-6).

To determine the importance of a given taxa predicted by RFC for a particular variable, we plotted the proportional abundance of the top ten taxa from the RFC analysis. This revealed that hunter-gatherers have higher abundances of *Bacteroidales, Prevotella stercorea, Succinivibrio, Phascolarctobacterium*, and *Treponema*; pastoralists have high abundances of *Odoribacter, Rikenellaceae, Bacteroides caccae*, and *Bacteroides ovatus*,

and agropastoralists have comparatively lower abundances for all of these taxa (Figure 4-4A).

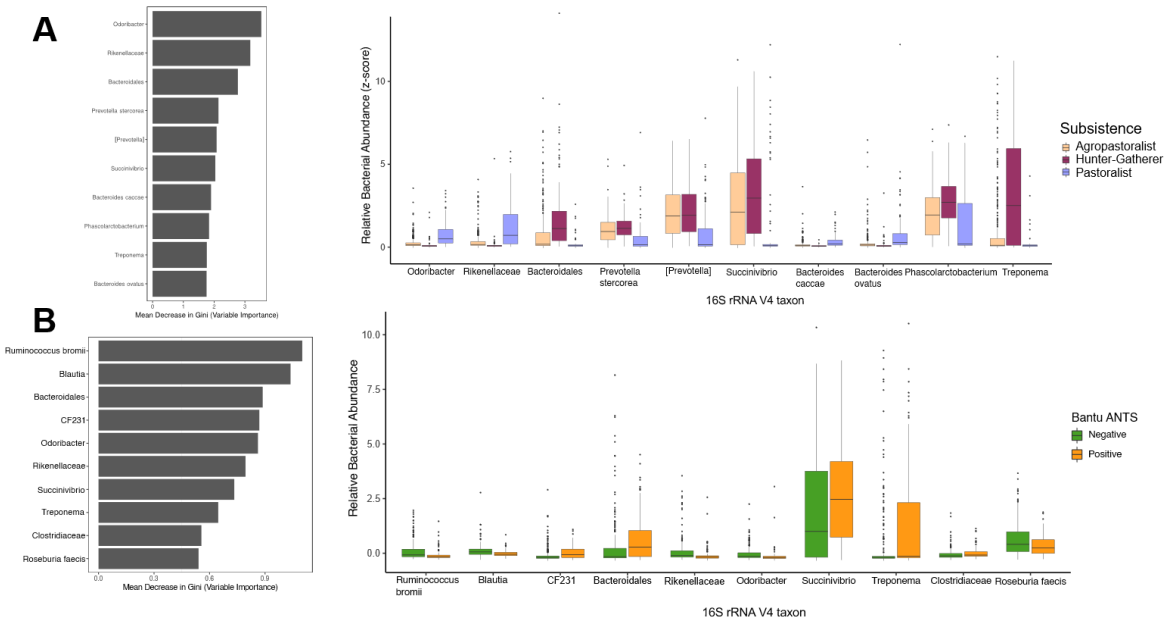


Fig 4-4. RFC on subsistence for all Cameroonians and ANTS infection in Bantu
A) Top ten most important taxa in predicting Cameroonian subsistence in 16S rRNA data, shown by model importance on left and relative bacterial abundance (z-scores) on the right.
B) Top ten most important taxa in predicting ANTS infection in Bantu agropastoralists only in 16S rRNA data, shown by model importance on left and relative bacterial abundance (z-scores) on the right.

An individual who was positive for any of the ANTS helminth parasites was considered "Positive" in an ANTS binary analysis. In our RFC analysis, *Bacteroidales, CF231*, *Treponema*, *Prevotella stercorea*, *Anaerovibrio,* and *Succinivibrio* increased with ANTS infection (Supp. Figure 4-6). In addition, we analyzed the Bantu agropastoralists separately, to avoid confounding with subsistence. In this subset of individuals, we found that the taxa most associated with infection were *CF231*, *Bacteroidales*, *Succinivibrio*, *Treponema*, and *Clostridiaceae* (Figure 4-4B). From these analyses, we find that the only bacteria predictive solely of ANTS infection and not simply predictive of subsistence

were *Ruminococcus bromii* (increased abundance in ANTS-negative individuals) and *CF231* (increased abundance in ANTS-positive individuals).

4.3.4. Relationships between ANTS infection and the microbiome using shotgun sequencing

We performed shotgun metagenomic sequencing on a subset of 175 Cameroonian fecal samples (94 Bantu, 37 Baka, 22 Bagyeli, 22 Fulani) to investigate the contribution of both bacterial and non-bacterial members of the microbiome to ANTS infection, as well as the associations between ANTS parasite positivity and gene function. These samples were selected to include ANTS-positive and negative individuals, diverse subsistence groups, and HIV-positive and negative individuals. Where possible, ANTS or HIV-positive subjects and controls were matched by sex, age, and sampling site (Supp. Table 4-1). The Cameroonian cohort was compared to 27 healthy U.S. human gut microbiome samples from the HMP cohort (Consortium, 2012; Human Microbiome Project Consortium, 2012).

Alpha diversity (Shannon Index) and evenness (Simpson's Index) results derived from shotgun metagenomic sequencing data were similar to those observed by 16S rRNA marker gene sequencing (Supp. Figure 4-7). The Fulani pastoralists were the only population that had significantly different microbial diversity from all other Cameroonian populations (1-way ANOVA with Tukey's honest significant difference test, adjusted p-values < 0.05), and had comparatively reduced alpha diversity by both metrics. However, samples from healthy U.S. urban individuals from the HMP cohort had significantly

lower alpha diversity than the Fulani population (p= 0.035 1-way ANOVA with Tukey's honest significant difference test on Simpson index).

Metagenomic reads were pre-processed using the Sunbeam pipeline (Clarke et al., 2019) and assigned to microbial taxa using several methods, including MetaPhlAn2 (Truong et al., 2015), KrakenUniq (Breitwieser, Baker, & Salzberg, 2018), and alignment to the Greengenes collection of 16S sequences (DeSantis et al., 2006; Wood & Salzberg, 2014). Across all three analyses, the top three bacterial genera were *Prevotella, Bacteroides* and *Faecalibacterium* (Figure 4-5A).

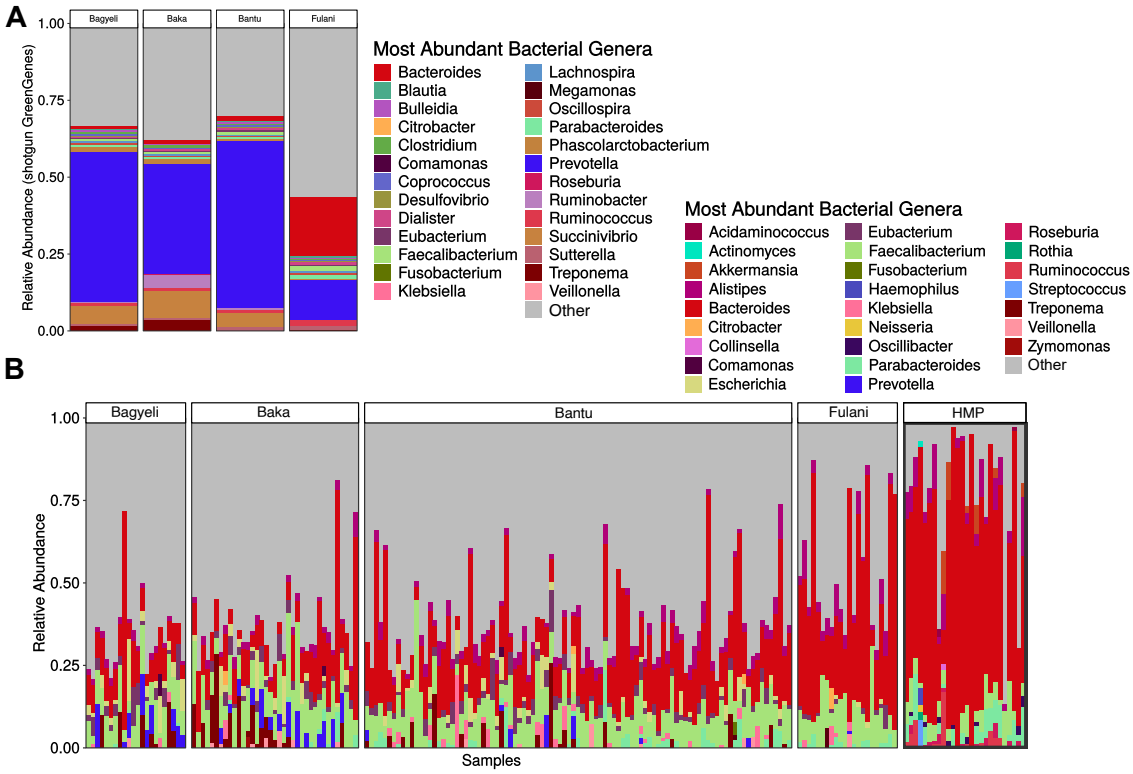


Figure 4-5. Shotgun Sequencing Composition of the Gut Microbiome
A) Fifteen most abundant bacterial genera per population, averaged across populations, studied in shotgun metagenomics analysis. B) Fifteen most abundant bacterial genera per population studied in shotgun metagenomics analysis. Individual vertical bars represent different samples. Human Microbiome Project samples shown for comparison at the far left show high levels of Bacteroides.

Shotgun metagenomic sequencing also showed that the Fulani pastoralists had higher levels of *Bacteroides* (average relative abundance = 34%) and lower levels of *Prevotella* (average relative abundance = 5.3%) than other Cameroonian populations, as was observed by 16S gene sequencing (Supp. Figure 4-8). Fulani and U.S. samples shared similarly high relative average bacterial abundances of *Bacteroides* (median 59% U.S., median 29% Fulani) and *Alistipes* (median 4% U.S., median 3% Fulani) compared to the other Cameroon populations (Figure 4-5B).

Using the KrakenUniq database we found that a larger fraction of metagenomic sequences remained unclassified in Cameroonian fecal samples compared to those from the HMP. Among the Cameroonian groups, the Fulani had the fewest unclassified sequences, which is likely attributable in part to their high *Bacteroides* abundance, members of which are well-studied and well-represented in genomic sequence databases (Pasolli et al., 2019). The Baka had the greatest number of reads unclassifiable at any level. Reads classified as viral, fungal, and parasitic-eukaryotic represented a small fraction of average total reads across samples (average 0.05%, max 1.5%). Although samples were not purified for virus-like-particles, we occasionally detected eukaryotic viruses including adenovirus and astrovirus (Supp. Figure 4-9). In particular, we noted that Cameroonians were more frequently positive for human mastadenovirus D than any other virus. Human mastadenovirus D species have associations with gastrointestinal, respiratory, and eye infections, have been found in river and drinking water in South Africa (van Heerden, Ehlers, Heim, & Grabow, 2005), and could be in Cameroonian water sources. However, we did not find significant differences between subsistence groups, sampling sites, or regions in frequency of human mastadenovirus D (Fishers's Exact test with Bonferroni test correction, p-values > 0.05). The detection of other eukaryotic viruses was too low to test for statistically significant differences between populations.

Metagenomic samples from four Cameroonian populations clustered by Bray-Curtis dissimilarities based on ANTS positivity (PERMANOVA p-values < 0.05) (Supp. Figure 4-10A). ANTS-positive Bantu agropastoralists were significantly different from ANTS-

negative Bantu agropastoralists (first principal component tested, Wilcoxon Rank-Sum test with continuity correction $p = 0.008$) (Supp. Figure 4-10B). This result emphasizes that the differences between ANTS positive and negative individuals in the entire Cameroonian cohort can be replicated in the Bantu alone, which is the only group in this study with adequate cases and healthy controls for within-group ANTS statistical tests. There was a significant correlation between molecular (qPCR cycle of threshold) and metagenomic detection (total k-mers) of *A. lumbricoides* ($p < 0.001$, Spearman's $\rho = -0.74$), *N. americanus* ($p < 0.001$, Spearman's $\rho = -0.63$ and *T. trichiura* ($p= 0.0008$, Spearman's $\rho = -0.36$) (Figure 4-6A).

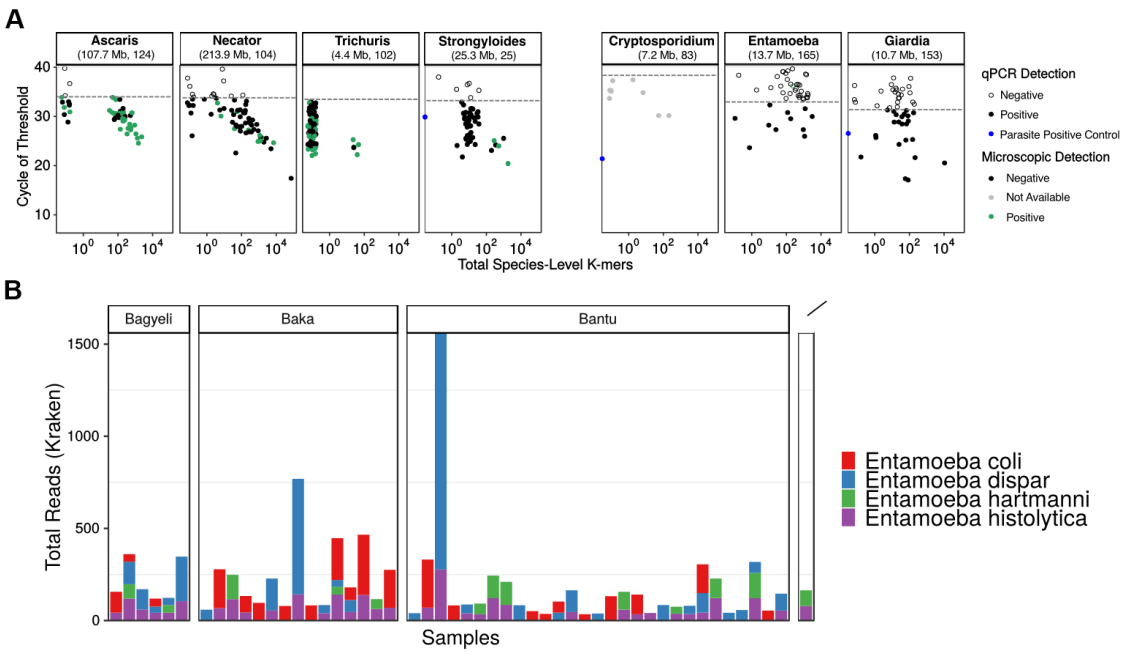


Figure 4-6. Comparisons of Parasite Detections using qPCR and shotgun sequencing A) Comparison of average Ct (Cycle threshold) values are shown along the y axis for ANTS (to the left of 4-6A), *Cryptosporidum* spp., *Entamoeba histolytica*, and *Giardia lamblia*. The x-axis corresponds to log-10 transformed total k-mers from shotgun metagenomic sequencing. A lower average Ct translates to an earlier cycle florescence threshold, and higher initial parasite genomic copy number. Lower Ct values have higher total k-mer counts and correspond to a higher copy of parasite genomes. The dashed line represents the highest standard. Blue dots along the y axis represent positive controls, which were not shotgun sequenced. Filled dots are samples that were positive in qPCR detection, and green filled dots were simultaneously positive in fecal microscopy. B) qPCR tested for *Entamoeba histolytica* only; however, given the amount of hits to qPCR negative targets, we looked at reads assigned at the species level to species of Entamoeba known to infect humans.

. For these parasites, higher parasite genome copies by qPCR were directly related to greater numbers of unique k-mers classified as those parasite species. We used unique k-mers (called by KrakenUniq) as a measure of genome coverage for this analysis rather than reads as unique k-mer counts are robust to potentially spurious read pileups. Furthermore, we saw a robust representation of large parasite genomes (*N. americanus*, *A. lumbricoides*) and limited detection of parasites with comparatively smaller genomes (*T. trichiura*, *Cryptosporidium* spp.), as expected. For *Cryptosporidium* spp. and *S.*

stercoralis, a positive trend was observed between molecular and metagenomics detection.

By incorporating multiple species of *Giardia* and *Entamoeba* parasites into the KrakenUniq database, we were able to identify infection with species other than those obtained in qPCR. All reads for *Giardia* mapped to sequences for *G. lamblia* or *Giardia intestinalis* (*G. intestinalis* and *G. lamblia* are names of the same organism). This difference in nomenclature for the same parasite explained the discrepancy in our early species-level shotgun taxonomic assignment. We identified different species of *Entamoeba* in shotgun analysis that were not detected by species-specific qPCR for *Entamoeba histolytica*. Of those known to infect humans (Fotedar et al., 2007), *Entamoeba dispar*, *Entamoeba coli*, and *Entamoeba hartmanni* were co-detected with *E. histolytica* in 33 samples (Figure 4-6B). With the exception of one Fulani individual who was positive for both *E. hartmanni* and *E. histolytica*, all other *Entamoeba* detections occurred in agropastoralists and hunter-gatherers. We again performed RFC analysis on the shotgun metagenomics data to validate 16S results and to test for additional associations between gut microbes and gut parasites not evident by 16S. Given the different species of *Entamoeba* detected in shotgun sequencing, three additional classification RFCs were run to test whether gut microbiota composition could predict positivity for commensal *Entamoeba* (*E. coli, E. dispar*, *E. hartmanni*), pathogenic *Entamoeba* (*E. histolytica*), or any *Entamoeba*, which had 69%, 54%, and 72% model accuracy, respectively. Of these, eight taxa had significantly different abundances between *Entamoeba* positive and negative categories (FDR-corrected Wilcoxon rank-sum test p-values < 0.05). Taxa that had significantly higher abundances in *Entamoeba* positive (Ent+) individuals were *Blastocystis hominis, Erysipelotrichaceae, Trueperella pyogenes*, and *Staphylococcus*

aureus. Taxa that had significantly higher abundances in *Entamoeba* negative (Ent-) individuals were *Flavobacterium magnum, Shigella dysenteriae (S. dysenteriae),* and *Anoxybacillus kamchatkensis*.

Blastocystis hominis (*B. hominis)* is a unicellular protozoan found in human large intestines and stool at rates higher than any other parasite in non-industrialized countries (Jantermtor et al., 2013; Stenzel & Boreham, 1996). Although *B. hominis* is usually considered a non-pathogenic commensal, it has been noted to associate with increased diversity of human gut bacteria of humans (Audebert et al., 2016). Members of the *Erysipelotrichaceae* family have been associated specifically with *Entamoeba* infection in western lowland gorillas (Vlčková, Pafčo, et al., 2018) and humans (Rosa et al., 2018). Both *S. aureus* (associated with Ent+) and *S. dysenteriae* (associated with Ent-) can induce changes in *E. histolytica* virulence and host response through modification of *E. histolytica* surface lectin expression, adhesion, cytotoxicity, and proteolysis (Bär, Phukan, Pinheiro, & Simoes-Barbosa, 2015). *Trueperella pyogenes* (Ent+) *Flavobacterium magnum (*Ent-), and *Anoxybacillus kamchatkensis* (Ent-) have not been associated before, to our knowledge, with *Entamoeba* infection status.

Subsistence RFCs using shotgun data were less accurate than 16S models (64% accuracy vs. 72%), but still detected three genus-level and species-level taxa that matched those implicated in the 16S analysis: *Bacteroidales bacterium CF, Phascolarctobacterium succinatutens, Treponema succinifaciens,* and *Bacteroides caccae* (Supp. Figure 4-11A). ANTS positivity was again a strong predictor of microbiome composition (77% accuracy) among tested metadata variables in shotgun RFC classification (Supp. Figure

4-11B), and had higher accuracy than the RFC for Bantu individuals who were ANTS positive (RFC 59% accuracy) (Supp. Figure 4-11C). Two taxa were identified that predicted ANTS positivity in both the Bantu and full Cameroonian cohorts, independent of subsistence: *Peptoclostridium acidaminophilum* and *Candidatus Azobacteroides pseudotrichonymphae*. Both taxa had significantly higher abundances in ANTS helminth positive individuals in the shotgun cohort and within Bantu-only (two-tailed Wilcoxon-rank sum test, p-values < 0.05).

Functional annotation of shotgun reads against Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000; Wixon & Kell, 2000) was performed using DIAMOND (Buchfink, Xie, & Huson, 2015). Pathway enrichment for Bantu individuals who were positive versus negative for ANTS, all Cameroon ANTS positive and negative individuals, and subsistence were assessed using Linear Discriminant Analysis Effect Size (LefSe) (Segata et al., 2011). Several KEGG classes were significantly differentially enriched amount the three subsistence groups (Figure 4-7A).

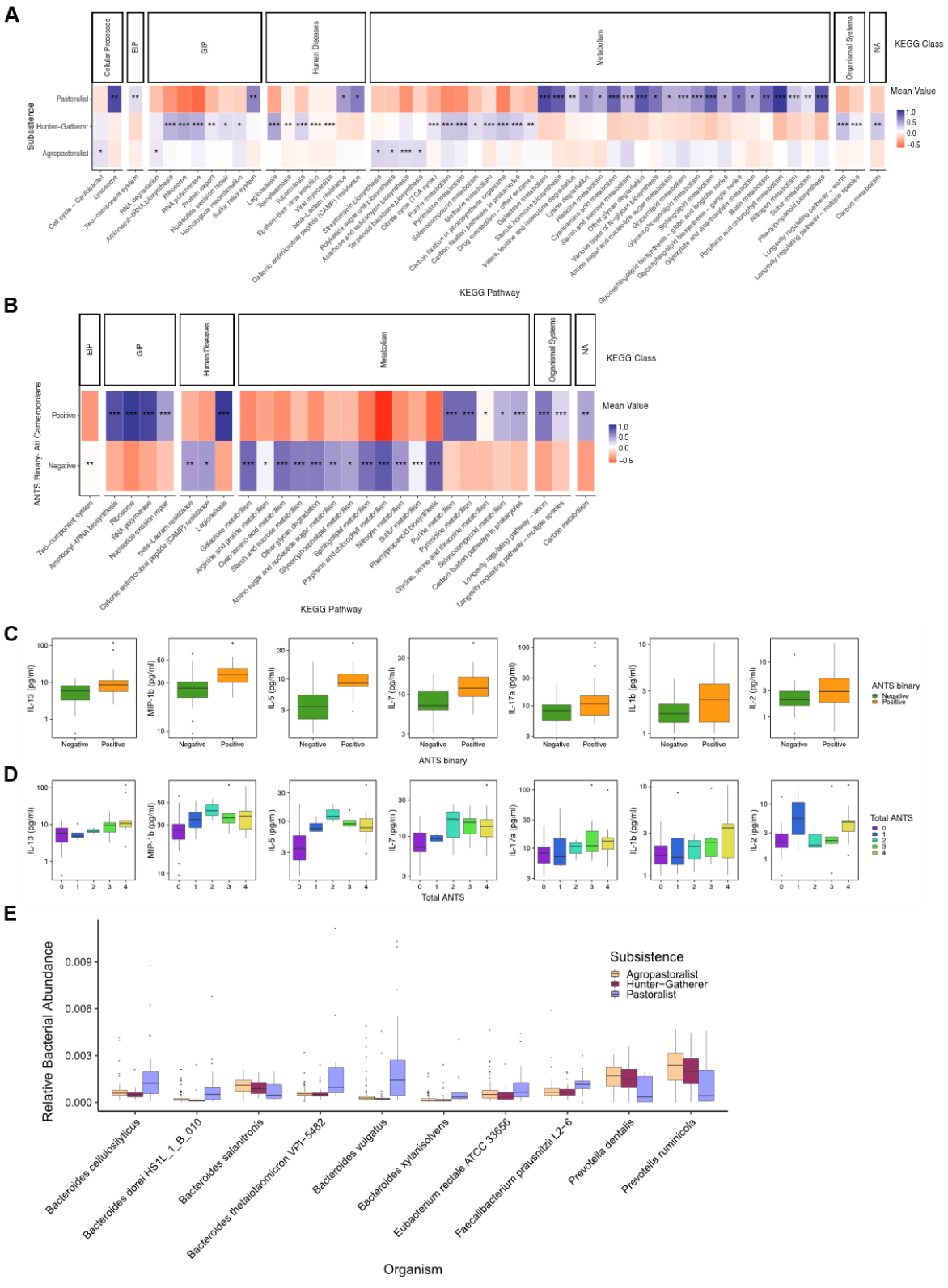

A
Cellular Processes
EIP
GIP
Human Diseases
Metabolism
Organismal Systems
NA
KEGG Class
Mean Value
Subsistence
Pastoralist
Hunter−Gatherer
Agropastoralist
KEGG Pathway
B
ANTS Binary- All Cameroonians
Positive
Negative
KEGG Class
Mean Value
KEGG Pathway
C
IL-13 (pg/ml)
MIP-1b (pg/ml)
IL-5 (pg/ml)
IL-7 (pg/ml)
IL-17a (pg/ml)
IL-1b (pg/ml)
IL-2 (pg/ml)
Negative
Positive
ANTS binary
ANTS binary
Negative
Positive
D
Total ANTS
E
Relative Bacterial Abundance
0.009
0.006
0.003
0.000
Subsistence
Agropastoralist
Hunter−Gatherer
Pastoralist
Bacteroides cellulosilyticus
Bacteroides dorei HS1L_1_B_010
Bacteroides salanitronis
Bacteroides thetaiotaomicron VPI−5482
Bacteroides vulgatus
Bacteroides xylanisolvens
Eubacterium rectale ATCC 33656
Faecalibacterium prausnitzii L2−6
Prevotella dentalis
Prevotella ruminicola
Organism

Figure 4-7. Functional analysis of gene content by pathway and association of ANTS parasites with cytokine levels
A) Functional analysis of gene content by pathway for ANTS using KEGG. B) Functional analysis of gene content by pathway for Cameroonian subsistence using KEGG. EIP: Environmental Information Processing. GIP: Genetic Information processing. C) Boxplots showing levels of all cytokines indicated as positively correlated in correlation analysis (Supp. Tables 8-9), binned by positive/negative status. D) Boxplots showing the same cytokines indicated as significantly correlated from C, but compared across counts of unique ANTS parasites (e.g., a dot in the "4" boxplot is an individual who was positive for all four ANTS parasites). Individuals are counted only once. E) Average abundance of bacteria from the KEGG galactose metabolism pathway in Fulani pastoralists with the LP and LNP phenotypes. The top ten most abundant bacteria produced from LefSe analysis for each phenotype are shown.

In agropastoralists, we found an enrichment of bacterial gene pathways involved in streptomycin biosynthesis, acarbose and validamycin biosynthesis, beta-Lactam resistance, and cationic antimicrobial peptide (CAMP) resistance. These gene pathways are all involved in the production of antibiotics or antibiotic resistance. Hunter-gatherers had an enrichment of genetic information processing pathways (e.g., Aminoacyl tRNA biosynthesis, RNA polymerase) and microbially-mediated disease pathways (e.g., tuberculosis, legionellosis), in addition to enrichment in methane, purine, and pyrimidine metabolism. Fulani had a bacterial gene enrichment for pathways in galactose, starch and sucrose, glycan, and lipid metabolism.

We again considered Bantu ANTS positive and negative individuals as a separate cohort in functional enrichment analysis given their balanced case/control ratios, but found that there were no significantly different pathways within Bantu individuals who were ANTS positive after multiple test correction. Across the entire Cameroon cohort, ANTS negative individuals tended to have pathway differentiation that closely followed the results for subsistence, and is likely a reflection of the comparatively larger populations

of ANTS infected individuals in hunting and gathering and agropastoralist groups versus pastoralists (Figure 4-7B). However, ANTS positive individuals had enrichment for genes that play a role in bacterial purine and pyrimidine metabolism, as well as longevity regulation. Purine and, less often, pyrimidine auxotrophic parasites must salvage these nucleotides from extraneous sources to synthesize DNA for their survival and proliferation (Gazanion & Vergnes, 2018), which may explain the enrichment in pathways that synthesize purines and pyrimidines. Purine-salvaging parasites that were found in the Cameroonians included the protozoans *Giardia* spp., *Plasmodia* spp., *Entamoeba* spp., and *Cryptosporidum* spp., as well as nematode *S. stercoralis* (Coghlan et al., 2019). When we looked at the bacterial contributing to enrichment of KEGG pathways in ANTS positive individuals from all Cameroonians, we identified a 100% overlap between these bacterial and the top ten taxa that were most predictive of ANTS helminth infection across all Cameroonians in shotgun RFC (Supp. Table 4-7). This result serves as an additional point of cross validation that these taxa and their functional pathways are significantly different for ANTS by predictive modeling and in LefSe analyses.

The concentrations of 19 cytokines across 72 Cameroonian plasma samples were measured and visualized in separate correlation plots by ANTS positivity ("ANTS binary"), ANTS total count (where count is 0-4), and HIV. ANTS positivity was significantly and positively associated with Th1 associated cytokines IL-7 and IL-2, Th2 cytokines IL-5 and IL-13, Th17 cytokine IL-17a, and proinflammatory cytokines MIP1b/CCL3 and IL-1b (Supp. Figure 4-12A; Figure 4-7C) (Spearman's correlation

coefficient, p-values < 0.01, Supp. Table 4-8). The count of ANTS parasites was positively associated with the same cytokines, except IL-2 (Supp. Figure 4-12B; Figure 4-7D) (Spearman's correlation coefficient, p-values < 0.01, Supp. Table 4-9). Levels of cytokines IL-13, IL-17a, and IL-1b increased with ANTS count, whereas cytokine values peaked at an ANTS count two for MIP-1b, IL-5, and IL-7. HIV had a positive but non-significant association with proinflammatory cytokine TNFα (Supp. Figure 4-12C, Supp. Table 4-10).

Regression-based RFC was performed on all cytokines in conjunction with 16S gene and shotgun metagenomic data to assess whether microbiome composition could be used to predict cytokine values (Supp. Table 4-6, Supp. Figure 4-12). Among all regression-based RFCs, and among all cytokines, IL-5 explained the most variation in 16S (31%) and shotgun (75%) microbiome data. The higher percentage of IL-5 explained variation in shotgun sequencing as compared to 16S gene sequencing is likely the result of the higher level of intentionally selected ANTS-positive individuals in the shotgun sequencing dataset, as IL-5 is positively associated with helminthiasis. IL-5 is essential in the development and recruitment of eosinophils to sites of infection and stimulates the production of anti-microbial peptides and mucus in the intestinal epithelium during helminthiasis (Annunziato, Romagnani, & Romagnani, 2015; Rosenberg, Dyer, & Foster, 2013). Of note, the *Bacteroidales* taxon explained the most IL-5 RFC model variation in shotgun analysis and the second most in the 16S gene analysis. In both cases, *Bacteroidales* was associated with a significant, positive relationship between *Bacteroidales* abundance and IL-5 level (Spearman's test for correlation in 16S p-value <

0.001; and ANOVA in shotgun p-value < 0.001). It was also the most explanatory taxon in highly positive ANTS, and ANTS infection ("Ants Binary") in shotgun and 16S RFCs. *Bacteroidales bacterium CF* had significantly higher abundance in ANTS positive versus ANTS negative individuals (Wilcoxon rank-sum test with continuity correction, p-value < 0.001) in shotgun analysis and in 16S gene (p-value < 0.001) analysis.

4.3.5. Gut microbiota are not associated with the lactase persistence phenotype in Cameroonians

The ability to breakdown lactose milk sugar in the small intestine past weaning and into adulthood is known as the lactase persistence (LP) phenotype (Arola, 1994). We predicted that hosts who are lactase non-persistent (LNP) would not be capable of metabolizing lactose in the small intestine and, instead, the lactose would be catabolized by bacteria containing the LAC operon in the distal colon. The LAC operon, which produces beta-galactosidase, cleaves the milk sugar lactose into galactose and glucose, and gut bacteria with this operon may play a role in the digestion of dairy products (Supp. Figure 4-13).

We administered a lactose-tolerance test (LTT) to 154 Cameroonians to test for association of the gut microbiome with the LP/LNP phenotype. Of the Cameroonians (see Methods, Supp. Table 4-1), 52 were LP and 102 were LNP. Although bacterial microbiome composition (16S) across the entire Cameroonian cohort did not differ by lactase persistence phenotype by weighted or unweighted PERMANOVA (Supp. Table 4-4, 4-5) and was not a strong predictor of microbiome composition in either 16S

(65.56% accuracy) or shotgun data (48.84% accuracy), we found an enrichment of genes in the galactose metabolism pathway for Fulani pastoralists (Figure 4-7A).

We traced these genes back to their bacteria of origin, which produced a list of over 1,619 unique bacteria in the Fulani with putative functional ability to catabolize galactose. We then filtered this list to inspect the top ten most abundant bacteria, and visualized their relative abundances by subsistence (Figure 4-7E). Although seven of these taxa were highest in Fulani pastoralists, the other three taxa, *Prevotella dentalis, Prevotella ruminicola,* and *Bacteroides salanitronis* all had higher abundances in agropastoralists and hunter gatherers than pastoralists. We then estimated the probability that high abundances of galactose associated bacteria identified here would produce an LP phenotype across the full Cameroonian cohort. Odds ratios were used to ascertain the likelihood that any of these ten taxa were higher or lower than their median abundances across the entire Cameroonian cohort in either LP or LNP individuals. We found that the taxa did not confer a higher chance of being LP or LNP across the entire cohort (Supp. Table 4-11). Although representing a small fraction of the total bacteria we identified as being able to breakdown lactose, we reasoned that they were most likely to have an effect on lactose phenotype since they were the most abundant However, given our results, we cannot rule out the possibility that all the galactose metabolizing bacteria may be together contributing to the catabolism of lactose in the distal colon, and could be considered in their entirety as a lactose-catabolizing community in future analyses. In summary, we can use shotgun sequencing data to infer dairy sugar catabolism in the colonic microbiota in

the Fulani, who have diets and subsistence practices consistent with high dairy consumption.

4.4. Discussion

Our results indicate that commensal microbiome compositions in Cameroonian populations are quite heterogeneous. These results parallel other studies of the gut microbiota of rural populations with traditional subsistence strategies (Arumugam et al., 2011; Ayeni et al., 2018; Clemente et al., 2015; De Filippo et al., 2010; Gomez et al., 2016; Morton et al., 2015; Obregon-Tito et al., 2015; Pasolli et al., 2019; Rampelli et al., 2015; Schnorr et al., 2014; Smits et al., 2017), while for the first time analyzing the impact of infection by multiple gastrointestinal parasites on the microbiota. We show that Cameroonians have higher amounts of *Prevotella* relative to *Bacteroides*-enriched U.S. samples. *Bacteroides* relative abundance is higher in Fulani than the other Cameroonian populations, which could be the byproduct of increased industrialization. Our ability to identify microbes from remote, rural settings is limited by the availability of appropriate reference microbial genomes. As a result, we have a larger fraction of "unknown" reads that could not be classified at any taxonomic level in Cameroonian samples as compared to U.S. samples. Our sample set also allows us to ask how the different baseline microbiotas in rural populations respond to colonization with multiple parasites.

Although our sampled populations live in rural areas, the Bantu and Fulani of Northwest Cameroon live in more developed villages with higher population density than the populations we sampled in South and East Cameroon. Furthermore, many individuals in

these two populations live in cooler, mountainous climates in the northwest. Since helminthiasis has been closely linked to moist, tropical environments, it's expected that we would have higher rates of parasites in southern and eastern populations. These factors may contribute to the higher rates of co-occurring fecal and blood parasites in the hunter-gatherers. Within Cameroonians, microbial alpha diversity positively correlated with gut parasite count. Increased gut bacterial diversity has been positively and negatively associated with intestinal helminth and protist infections (Leung et al., 2018). Here, we show that increasing parasite count correlated with higher bacterial diversity. We speculate that this higher diversity could be associated with disease morbidity, where factors such as parasite-induced rapid peristalsis and epithelial shedding could produce an optimal environment for opportunistic microbial colonization.

Recently, Morton et al. (2015) characterized the bacterial microbiome of rural populations in southwest Cameroon that were highly parasitized by protists and helminths through 16S rRNA gene sequencing. Morton et al. (2015) found that *E. histolytica* colonization was predictive of microbiome diversity and composition (79% accuracy). Here, we found that when *A. lumbricoides, N. americanus, T. trichiura,* and *S. stercoralis* helminths are considered as co-occurring morbidities, the cumulative effects of ANTS helminth infection are highly predictive of microbiome composition (81% accuracy). The use of molecular techniques in this study enabled the quantification of species that are not morphologically distinguishable by light microscopy (*Cryptosporidium* spp., multiple *Entamoeba* species). Although we did not have enough qPCR positive *E. histolytica* samples for species-specific RFCs, shotgun sequencing

revealed that several Cameroonians were infected with multiple types of commensal and pathogenic *Entamoeba*. Coinfection with *Entamoeba* commensal and pathogenic species has been previously documented in human and nonhuman primate hosts in east Cameroon (Vlčková, Kreisinger, et al., 2018). The degree to which human commensal and pathogenic enteric *Entamoeba* interact, enhance, or inhibit one another remains an open area of investigation.

Multiple RFC models testing different categories of ANTS parasite infection and IL-5 cytokine response indicated that *Bacteroidales* was an important predictive taxon. *Bacteroidales* was consistently found at elevated abundances in ANTS positive individuals in comparison to ANTS negative in this study. Previously, *Bacteroidales* has been found in lower abundances in the guts of humans infected with *Entamoeba histolytica* (Morton et al., 2015). In murine models, infection with helminths led to the reduction of *Bacteroidales* and the concurrent expansion of *Clostridiales* communities (Ramanan et al., 2016), which was hypothesized to stimulate an anti-inflammatory response (increased IL-5 and IL-13) in the host. In this study, rural Cameroonians with high abundances of *Bacteroidales* were a strong predictor of helminthiasis and Type-2 immune response (IL-5) in our dataset. This observation is consistent with similar *Bacteroidales* expansions detected in the gut microbiomes of rural Malaysians infected with *T. trichiura* (Lee et al., 2014); however, we cannot rule out that *Bacteroidales* abundances found here may be confounded with subsistence practice. Rural populations may have different microbial responses to particular types of parasite infections, which may be shaped in part by subsistence, host genetics, geography, or other factors. Whether

or not any of these differences may contribute to host anti-inflammatory properties and parasite clearance in the absence of antihelminthic drugs remains to be investigated.

Furthermore, RFC results suggested that four taxa, *Ruminococcus bromii, CF231, Peptoclostridium acidaminophilum* and *Candidatus Azobacteroides pseudotrichonymphae* may have roles in ANTS infection exclusive of subsistence. *CF231* is a common occupant of ruminant guts (Wang et al., 2016) that has occasionally been found in humans (Yun et al., 2017) but has not, to our knowledge, been associated with human fecal parasite infection. *Ruminococcus bromii* has functions in degrading foods high in resistant starch and has been associated with fishing subsistence in coastal Cameroonians (Morton et al., 2015). In our study we found that individuals with this bacterium were less likely to be infected by ANTS when controlling for subsistence. Diets rich in starch-resistant foods have been linked to reductions in gastrointestinal inflammation (Lockyer & Nugent, 2017), making it possible that *R. bromii* could have a protective effect against helminthic disease, or at least in alleviating helminth-associated inflammation. *Peptoclostridium acidaminophilum* (previously known as "*Eubacterium acidaminophilum*" (Galperin, Brover, Tolstoy, & Yutin, 2016) is a versatile, amino-acid degrading anaerobe that has not been associated, to our knowledge, with helminthiasis in prior research. However, Hadza hunter-gatherers were previously described to have an enrichment of KEGG Orthologous genes involved in amino acid metabolism, and greater functional potential to metabolize proline, serine, glycine, and threonine (Schnorr et al., 2014, p. 201). *Candidatus Azobacteroides pseudotrichonymphae* is a termite endosymbiont (Pramono et al., 2017) and unclassified taxa in the carbohydrate-

metabolizing *Ruminococcaceae* family and nonpathogenic species in the genus *Treponema* (both of which are present in our data) are considered common occupants of termite guts (Angelakis et al., 2018). Termites are a substantial component of hunter-gatherer and agropastoralist diets in many parts of Cameroon (Sato, Kawamura, Hayashi, Inai, & Yamauchi, 2012; Tamesse, Kekeunou, Tchouamou, & Meupia, 2018). *Treponema* is also a common constituent of healthy nonhuman primate guts (Clayton et al., 2018; McKenna et al., 2008). Termite consumption could be more common in individuals that are ANTS positive, which could be affected by bioavailability of termites, subject to climate and location (i.e., more tropical locations, which are correlated with a higher infectious disease burden of helminths). Whether the presence of these taxa has any effect on susceptibility or response to ANTS helminth infection remains an open question.

In addition, we studied HIV+ subjects who had no obvious symptoms and found no detectable microbial community alterations (Supp. Figure 4-15 and 4-16), paralleling some but not all studies of lentiviral infection and the gut microbiome. We did find a modest positive correlation between TNFα and HIV infection. The TNFα/TNFR pathway has been established as a component of immune activation and the development of viral reservoirs during HIV infection (Pasquereau, Kumar, & Herbein, 2017).

Given the dairy-rich diet of Fulani pastoralists, and a reported 50% prevalence of clinical lactose tolerance (Enattah et al., 2007; Lokki et al., 2011; Tishkoff, Reed, et al., 2007), we had expected that many of the Fulani would produce the lactase enzyme, which cleaves lactose into glucose and galactose. If the host can successfully take up lactose in

the small intestine then less is passed to the distal colon, which would hypothetically result in lower abundances of gut bacteria capable of catabolizing lactose in lactase-persistent hosts as compared to lactase non-persistent hosts. When we tested for association of the gut microbiome with the lactase persistence phenotype, we did not detect a significant correlation with the gut microbes between LP and LNP individuals or with the taxon *Bifidobacterium*, which has been extensively associated with dairy catabolism in populations with majority European ancestry (Goodrich et al., 2017).

We found that pastoralists had higher levels of galactose-metabolizing microbial genes compared to the other Cameroonians. The abundances of the top ten bacterial taxa implicated as most correlated with Fulani galactose metabolism had variable abundances among the subsistence groups. Notably, galactose metabolizing *Prevotella dentalis* and *Prevotella ruminicola* had comparatively higher abundances in agropastoralists and hunter-gatherers across the entire Cameroonian cohort. Consumption of short-chain galactooligosaccharides, has been associated with increased *Prevotella* abundance and attenuated lactose tolerance in the guts of lactose-intolerant individuals (Azcarate-Peril et al., 2017). As such, *Prevotella* species may provide lactase non-producers with an enhanced ability to degrade milk sugars. Overall, these results indicate that several individual bacteria may be capable of catabolizing lactose sugar and have roles in galactose catabolism within Fulani pastoralist gut microbiomes.

4.5. Conclusions

This study represents the largest work to date on the correlations between polyparasitemia (fecal and blood) and gut microbiota in sub-Saharan Africans. We establish that cooccurrence of gut parasites is significantly associated with microbial community structure in the gut, and identify putative taxa associated with subsistence type, cytokine response, and the ANTS helminth infection. Further studies would benefit from longitudinally sampling populations and integrating individualized dietary information to establish baseline, to distinguish healthy host microbiome structure from parasitized states, and to test for association of microbial diversity with seasonality. The gut microbiota are an intriguing potential therapeutic target in the treatment or prevention of helminthiasis, motivating further investigation into the mechanisms behind parasite-host-microbiota interactions.

4.6. Methods

4.6.1. Subject Details

Written, informed consent was obtained from all study participants and research/ethics approval was obtained from the following institutions prior to the start of sample collection: Institutional Review Board of the University of Pennsylvania, the Cameroonian National Ethics Committee and the Cameroonian Ministry of Public Health. All subjects provided written informed consent for the collection and analysis of samples. All samples were coded with an alphanumeric identifier to protect participant confidentiality.

Cameroon samples were collected from nine villages in the Northwest (Ntambang, Sabga), South (Bidou I, Ndtoua), and East (Nkolbikong, Missoume, Njibot, Aviation, Bosquet) regions of Cameroon (Figire 4-1A), all of which represented rural communities. The villages in the East and South Administrative regions are located in densely forested areas with primarily tropical monsoon and rainforest climates (Köppen-Geiger climate classification), while the villages in the northwest are primarily in tropical savanna climates (https://en.climate-data.org/location/2905/). The traditional wet season for Cameroon is April through September, and the traditional dry season runs October through March. The Cameroon populations were sampled between January to July of 2015. All populations sampled in this study speak languages in the Niger-Kordofanian family. Ethnicity, sample sizes, sampling coordinates, and subsistence classifications are listed in Supp. Table 4-1.

DNA extracted from fecal samples from 37 healthy, omnivorous, U.S. participants in the greater Philadelphia, was used here for comparative purposes in the 16S analyses. Eleven of the U.S. participants self-described their ethnicity as African-American, and 26 self-described as European-American. These samples were collected for a prior research study, the details of which can be found in Wu et al., 2011. Data for age, sex, height, weight, location, and BMI, were included in diversity metric analyses (Supp. Table 4-1).

4.6.2. Sample collection and storage

Fecal samples were obtained from asymptomatic subjects with no signs of clinical illness and who self-reported as not pregnant. Participants produced a fecal sample in a sterile

plastic container that was immediately returned to researchers at the field site. A midsection sample of stool (~5 g aliquots) was harvested into a 5 ml container and immediately frozen in liquid nitrogen. Samples were stored at -80°C before transportation to the US in dry ice, where it was again stored at -80°C until extraction.

For 524 of these individuals, contemporaneous plasma samples were also collected. Blood was drawn into 10ml capacity BD Vacutainers containing EDTA, and small drops of blood were taken from this tube to measure white blood cell count (HemoCue WBC analyzer and HemoCue WBC cuvettes), and to make thick and thin blood smears on slides for malaria and filarial parasite testing. Relative percentages of lymphocytes, monocytes, eosinophils, basophiles, and neutrophiles were measured for 570 individuals. Following this step, the tube of blood was immediately spun down and plasma was processed through a Leukolock kit (Ambion Inc.). Plasma was aliquoted into 0.5ml Eppendorf SafeLock tubes and frozen in liquid nitrogen. The plasma was frozen at -20°C and all samples were analyzed simultaneously.

4.6.3. Fecal sample processing and DNA sequencing

4.6.3.1. Fecal DNA Extraction

Cameroonian and U.S. fecal samples were processed with the same laboratory and computational pipelines for extraction and 16S analysis. Total DNA from fecal materials was extracted from ~220mg aliquots using a PSP Spin Stool DNA Plus Kit (Stratec Biomedical, Germany) with a modified bead-beating method (Salonen et al., 2010). PCR and extraction blanks were used to control for reagent and environmental contamination,

and all extractions were conducted in a laminar flow hood. Eluted DNA was quantified by fluorometry and stored at -20°C.

4.6.3.2. Bacterial 16S rRNA amplicon sequencing

PCR reactions were performed on extracted fecal DNA in triplicate using Accuprime Pfx Supermix (Invitrogen) and barcoded composite primers with Illumina adapters to amplify the V4 region of the bacterial 16S rDNA genome following the methods of Kozich et al. (2013) on a GeneAmp 9700 PCR System. Sequences of DNA primers used in this study are reported in Supp. Table 1. PCR conditions were as follows: 95°C for 2 min, followed by 30 cycles of 95°C for 20 sec., 55°C for 15 sec., 72°C for 5 min., and then a final elongation step at 72°C for 10 min. A gene block mock community of eight archaeal species not normally detected in experimental data was used as a positive control following Kim et al. (2017) (Supp. Table 4-12). Samples containing the resulting ~250 bp products were pooled, and a subset were visualized by gel electrophoresis on a 1% agarose gel. Library clean-up was performed using SequalPrep Plate Normalization Kits (Invitrogen), and average library fragment size was checked on a subset of samples using a Tapestation d1000 ScreenTape System (Agilent). Libraries were quantified using Qubit dsDNA HS Assays (Thermo Fisher Scientific) and pooled in equal amounts. Libraries were sequenced on an Illumina MiSeq across 4 runs using 2 x 250 bp cycles in the Bushman Lab. Sequence data are deposited under project accession PRJNA547591 in the NCBI Sequence Read Archive; sample details are in Supp. Table 4-1.

4.6.3.3. Shotgun metagenomic sequencing

A total of 178 fecal DNA sample aliquots and controls were normalized to 0.2 ng/ul DNA and 1 ng of DNA per sample was used as input for the Nextera XT DNA Sample Prep Kit (Illumina Inc.) and manufacturer protocols. PCR amplification using unique combinations of barcoded primers was performed on a GeneAmp 9700 PCR System, and short DNA fragments were removed using AMPure XP bead purification. Library fragment size was visualized on a Tapestation d1000 ScreenTape System (Agilent) and libraries were quantified using PicoGreen before being pooled in equimolar ratios for sequencing. Three extraction negative controls (denoted "EB" in the metadata) and two library negative controls ("Lib Neg") were included on the run. The pooled library was subjected to a second round of quantification on a BioAnalyzer 2100 (Agilent), followed by a MiSeq Nano sequencing run for quality control. After this, the pooled library was diluted in hybridization buffer, heat denatured, and paired-end sequenced on an Illumina HiSeq 2500 using V4 reagents in the Penn CHoP Microbiome Core.

4.6.4. Parasite and HIV testing

4.6.4.1. Microscopy

Stool samples were examined in the field for parasite presence using wet-mount fecal microscopy. Samples were examined with and without iodine staining and visualized with standard light microscopy to identify visible gastrointestinal parasites or parasite ova, including hookworm (species indeterminate with light microscopy), amebiasis (*Entamoeba* spp.), giant roundworm (*Ascaris lumbricoides*), human whipworm (*Trichuris trichiura*), giardia (*Giardia* spp.), and human roundworm (*Strongyloides*

stercoralis). Thick and thin blood smear slides were prepared with Giemsa staining to identify blood parasites in the field, including plasmodia (*Plasmodium* spp.) and filaria (*Microfilaria loa loa, Microfilaria* spp., *Mansonella perstans, Wuchereria bancrofti*). Details on microscopy positivity are in Supp. Table 4-1.

4.6.4.2. Quantitative PCR (qPCR)

DNA oligonucleotide sequences used in qPCRs are listed in Supp. Table 4-13. A gBlocks gene fragment (IDT) containing parasite target sequences was synthesized and cloned into a TOPO cloning vector, transformed into TOP10 competent E. coli cells and purified with a Qiaprep Spin Miniprep Kit (Qiagen). Purified plasmid DNA was quantified by Picogreen and the sequence was validated with Sanger sequencing. Plasmids were diluted to a known concentration and serial 1:5 dilutions were performed to generate a 9-point standard curve. Unknown samples were compared against this standard curve for quantification. Positive control DNA was extracted from three parasite samples: *Cryptosporidium parvum* from infected mouse stool sample, and *Giardia lamblia* and *Strongyloides stercoralis* from infected canine stool samples using the same methods as human stool DNA extraction. Wells with no template were used as negative controls, and all controls and standards were tested in duplicate. Species-specific primers and probes used in Mejia et al. (2013) were used to assay parasite genome copy number for *Ascaris lumbricoides, Necator americanus, Ancylostoma duodenale, Giardia lamblia, Entamoeba histolytica, Trichuris trichiura,* and *Strongyloides stercoralis* parasites. The pan-*Cryptosporidium* spp. qPCR uses primers and probes from Jothikumar et al., 2008 which tests for ten *Cryptosporidium* species: *C. hominis, C. parvum, C. canis, C. felis, C.*

parvum-like (from lemurs), *C. muris, C. andersoni, C. baileyi, C. wrairi,* and *C. serpentis*. All qPCRs were conducted on individual parasites using 384-well MicroAmp EnduraPlate Optical 384-Well Clear Reaction Plates (Applied Biosystems) in triplicate on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems). The total volume per reaction was 7 ul, consisting of 3.5 ul of Taqman Fast Advanced Master Mix (Applied Biosystems), 2 ul of template DNA and 1.44 ul of species-specific primers (final concentration of 900 nM) and probes (final concentration of 250 nM), and 0.06 Sigma water (Sigma-Aldrich). qPCRs were run with default parameters and 40 cycles. In this study, we report quantification cycle threshold (Ct), which correspond to the PCR cycle values measuring when fluorescence from template amplification exceeds background fluorescence and is an inverse measure of nucleic acid. At least two of the three replicates had to fluoresce within the standard range for the sample to be positive. We had no samples that were positive for *Ancylostoma duodenale,* and thus this parasite was removed from all downstream analyses.

4.6.4.3. HIV testing

For all participants who produced a plasma sample, simultaneous qualitative testing for Human Immunodeficiency Virus (HIV) p24 antigen and antibodies to HIV Type 1 (HIV-1 Groups M and O) and HIV Type 2 (HIV-2) was done using a GS HIV Combo Ag/Ab EIA immunoassay (BioRad). Testing was performed on 75 ul per sample of thawed plasma according to manufacturer's instructions. Results were read on a SpectraMax 190 absorbance microplate reader (Molecular Devices). In addition to positive controls from the GS HIV Combo Ag/Ab EIA kit, human serum from an anonymous, seropositive

donor from the U.S. who had not yet been treated with antiretroviral drugs was used on every test plate.

4.6.5. Serum cytokine measurements

We measured 21 cytokines from plasma using a high-sensitivity multiplex cytokine panel (Milliplex MAP Human High Sensitivity T Cell Magnetic Bead Panel, 21-Plex). They included Fractalkine/CX3CL1, granulocyte macrophage colony-stimulating factor (GM-CSF), interferon-gamma (IFNγ), interleukin (IL) 1β, IL-2, Il-4, IL-5, IL-6, IL-7, Il-8/CXCL8, IL-10, IL-12 (p70), IL-13, IL-17A/CTLA8, IL-21, IL-23, I-TAC/CCL11, MIP-1α/CCL3, MIP-1β/CCL4, MIP-3α/CCL20, and TNFα (tumor necrosis factor α). The panel was run on a Bio-Plex 200 machine using the manufacturer's protocols (Millipore Sigma). Cytokine concentrations were determined using standard curves, with the limits of detection for analytes reported in Sup. Table 4-1. A total of 72 Cameroonian samples were analyzed in two batches, with high and low cytokine-specific controls used across both batches. Measurements for each sample and standard curve were performed in duplicate, with the average of the two measurements reported. For two cytokines (MIP-1α/CCL3 and IL-2), we had an insufficient amount of non-NA values to conduct statistical tests, and we removed these from further analysis.

4.6.6. Lactose tolerance test and lactase persistence association test

4.6.6.1. Lactose tolerance test

To test for association of the gut microbiome with the lactase persistence phenotype and genetic variants that confer lactase persistence, 154 individuals from the Cameroonian

cohort were given a lactose tolerance test (LTT) (Ranciaro et al., in preparation). Participants fasted overnight and had baseline glucose measured before beginning the test using either a Code Free glucometer with SD CodeFree strips or an Accu-Chek Active glucometer with Accu-Chek Active strips. Exclusion criteria included having a baseline glucose outside of 60-100 mg/dl and diabetes. Participants drank a 50 g lactose powder solution (QuinTron USA) dissolved in 250 ml water which was equivalent to ~1-2 liters of cow's milk (Arola, 1994). Blood glucose was measured in 20-minute intervals over the next hour.

4.6.6.2. Lactase persistence association test

Glucose values were first adjusted to correct for test strip error using the regression equation $y = 0.985x - 7.5$, where x is the measured glucose value. The maximum rise in glucose level was ascertained by comparing observed glucose values against the baseline value and used to classify lactase phenotype. Individuals were classified as either lactase persistent (LP) (rise in blood glucose >1.7 mmol/l), lactase non-persistent (LNP) (rise in blood glucose <1.1 mmol/l), or lactase intermediate persistent (LIP) (rise in blood glucose between 1.1 and 1.7 mmol/l). We tested for enrichment of counts of lactase persistence associated alleles in either LIP or LP individuals (denoted as the lactose tolerant trait, also referred to as lactose digesters) relative to LNP individuals (lactose non-digesters). Of 154 participants, 102 were LNP, 21 were LP, and 31 were LIP. The latter two groups are both considered "Positive" in LP binary analysis, where samples are considered either lactase persistent (n = 52) or non-persistent (n = 102).

4.6.7. Anthropometry

4.6.7.1. Height

Height was measured using a Shorrboard Stadiometer (www.shorrproductions.com), with the individual in an erect position with the Frankfurt plane as horizontal as possible. Height was measured with shoes if worn (this was noted in the anthropometry form). Height was recorded in centimeters to the nearest millimeter.

4.6.7.2. Temperature

Temperature was measured in triplicate on a non-contact infrared thermometer.

4.6.7.3. Weight

Weight was measured using a set of Seca 876 scales and recorded in kilograms to one decimal place. Weight was measured with shoes if worn. Care was taken to ensure that the scales were firmly seated and level.

4.6.8. Quantification and statistical analysis

4.6.8.1. ASV inference

The V4 region of the bacterial 16S rRNA gene was sequenced on the Illumina MiSeq platform across 4 MiSeq runs. FASTQ files were generated from raw BCL files using "configureBclToFastq.pl" (Illumina Inc.) and paired-ends were assembled using the QIIME2 pipeline. All sequences went through quality-filtering, demultiplexing, chimera removal, denoising, and merging using the demux and DADA2 plugins with default settings. DADA2 produces an amplicon sequencing variant (ASV) table that can resolve unique sequences down to single-nucleotide differences and attaches biological meaning

to sequences independent of a reference database. All ASV feature tables were then merged (https://github.com/marubel/R-ubelMisc). We used a classifier that was pretrained on the V4 region targeted by the 515F and 806R primer sets (Caporaso, Kuczynski, et al., 2010; Caporaso et al., 2011) with 99% OTU sequence similarity using the most recent version of the GreenGenes (http://greengenes.secondgenome.com/) database. Sequences classified as mitochondria and chloroplasts were removed. To conduct phylogenetic analyses of microbiome sequences, sequences were aligned with MAFFT and a phylogenetic tree was produced with FastTree2 using default settings. Sequences derived from plastids and mitochondria were removed. A gene block mock community of eight archaeal species not normally detected in experimental data was used as a positive control across runs (Supp. Table 4-12), following the methods used in Kim et al., (2017). The four gene block controls, 12 negative extraction and PCR controls were dropped from further analysis. This produced a total of 14,138 ASVs.

4.6.8.2. Shotgun metagenomic sequencing processing and analysis

Illumina Basespace output metrics from the shotgun metagenomics run are available in Supp. Tables 4-14:4-16. In brief, there were 1,820,262,487 raw reads and 1,820,262,480 reads that passed the Illumina chastity filter. The Illumina chastity filter measures the ratio of the intensity base call divided by the sum of the brightest and second brightest intensity base calls. Raw shotgun metagenomic data files from were de-multiplexed and converted from BCL to FASTQ using bcl2fastq (Illumina), which drops unassigned reads, including those mapping to PhiX, a common sequencing control. Demultiplexed FASTQ files were analyzed using the Sunbeam pipeline (Clarke et al., 2019), as detailed

in (https://github.com/sunbeam-labs/sunbeam/). In short, quality filtering was done using default settings of Trimmomatic (reads below 36 bases, trailing or leading bases with quality scores below three, and base reads scanned in a 4-nt sliding window with average quality/base < 15 were dropped) (Bolger, Lohse, & Usadel, 2014) and adapters were trimmed from sequences with Cutadapt (fwd_adapters: ['GTTTCCCAGTCACGATC', 'GTTTCCCAGTCACGATCNNNNNNNNNGTTTCCCAGTCACGATC']

rev_adapters: ['GTTTCCCAGTCACGATC', 'GTTTCCCAGTCACGATCNNNNNNNNNGTTTCCCAGTCACGATC'] (M. Martin, 2011) software. This effectively dropped the two library negatives and the three extraction blanks from further shotgun analyses. FastQC (Babraham Bioinformatics) was used to assess read quality on read pairs surviving quality filtering (Supp. Table 4-15). Low complexity sequences were masked using (https://github.com/eclarke/komplexity) with a normalized complexity score of < 0.55. For k=4, this scores that across a 64-120 bp region, the sequence is strongly suggestive of being low-complexity, repetitive sequence, and thus is unlikely to be informative. Reads that mapped to a human reference sequence (Genome Reference Consortium Human Build 38, GRCh38) were identified using bwa (H. Li & Durbin, 2009). Samples reads with > 60% of the read fraction mapping to GRCh38 or with a percent identity >50% and were removed. Per sample nonhost (microbial) reads can be found in Supp. Table 16. Output from the Sunbeam quality control was inspected manually using the sbx_report extension (https://github.com/sunbeam-labs/sbx_report). This produced a total amount of 1.65 billion host-filtered, quality-controlled reads (controls not included). This amounts to a

median of 8.5 million reads per sample, or an average of 9.3 million reads. KrakenUniq (Breitwieser et al., 2018) was used to classify human-filtered, quality-controlled reads using the Sunbeam extension sbx_kraken_uniq (https://github.com/ArwaAbbas/sbx_kraken_uniq) on a low-complexity masked database of bacterial, archaeal, viral, fungal and protozoal sequences from NCBI nt (downloaded 13 December 2018). Classifications reported at genus and species level are reported as relative abundances and were further filtered based on a meeting a threshold of number of reads and read:k-mer ratio, as described in the figure legends. Shotgun metagenomic reads were also classified using two alternate methods for comparison: MetaPhlAn2 (Truong et al., 2015) on the MetaPhlAn2 mpa_v20 database using the sbx_metaphlan Sunbeam extension (https://github.com/sunbeam-labs/sbx_metaphlan/), and the GreenGenes 16S database using Kraken2 (https://ccb.jhu.edu/software/kraken2/). As expected given the differences in the database used, and known 16S primer biases, the relative proportions of bacterial genera classified by amplicon and shotgun sequencing did not always correspond. In a comparison across MetaPhlAn2, KrakenUniq, and 16S Greengenes classification for both V4 and shotgun data, comparisons between V4 and shotgun data annotated against the same Greengenes Database most closely paralleled each other (Supp. Figure 4-8). Second to that, the KrakenUniq database showed the least divergence between V4 and shotgun taxonomic identification. We note that some relevant genera were highly divergent in KrakenUniq to V4 comparison, including *Klebsiella* and *Eubacterium*, which were absent from V4 datasets. This could be the result of primer or database bias.

We note that average 16S copy number across *Prevotella* and *Bacteroides* genomes varies (average 16S copies across 24 species of *Prevotella* = 4, average across 24 species *Bacteroides* = 5.3) (https://rrndb.umms.med.umich.edu/genomes/). In our shotgun metagenomics data, the average size of *Bacteroides* genomes was 5.3 Mbp and the average size of *Prevotella* genomes was 2.7 Mbp. Larger genome size and more 16S copy numbers in *Bacteroides* could account for some of the variation we see in higher relative abundances of this taxon in 16S versus shotgun sequencing compared to *Prevotella*. Quality-controlled reads were aligned to the KEGG database (Abbas et al., 2019; Kanehisa & Goto, 2000; Wixon & Kell, 2000) using DIAMOND (Buchfink et al., 2015), using an e-value cutoff $1x10^{-6}$. The resulting KO numbers were mapped to the associated pathway, module, and enzyme identifiers (https://github.com/marubel/kegg-r-ator). Where a single KO mapped to multiple pathways, enzymes, or modules, weighted counts were used such that each KO contributed a single count equally distributed across all pathways, enzymes, or modules mapped to it.

To detect helminths in shotgun metagenomic sequences, human-filtered, quality-controlled reads were aligned to nine representative genomes downloaded from NCBI Genome (*Trichuris trichiura*, *Ancylostoma duodenale*, *Strongyloides stercoralis*, *Ascaris lumbricoides*, *Necator americanus*, *Entamoeba histolytica*, *Giardia lamblia*, *Cryptosporidium hominis* and *Cryptosporidium parvum*) or WormBase ParaSite (*Ascaris lumbricoides*). Read alignments were performed using hisss (https://github.com/louiejtaylor/hisss) (Abbas et al., 2019) using the following modifications to Bowtie2 (-end-to-end --very fast).

Results from all analyses were visualized in R (Ihaka & Gentleman, 1996) using packages tidyverse (Wickham, 2017), taxonomizr (Sherrill-Mix, 2018), magrittr (Bache & Wickham, 2014), reshape2 (Wickham, 2007), ggplot2 (Wickham, 2016), vegan (Oksanen et al., 2013), and ape (Paradis, Claude, & Strimmer, 2004). Pathway, module, and enzyme differential enrichment were calculated using LefSe (Segata et al., 2011), which produces absolute values of log10 transformed LDA scores as effect sizes for a given taxa/group. Code used to generate LefSe metrics and heatmaps can be found at https://github.com/ressy/LEfSe. FDR correction was applied to all LefSe results.

4.6.8.3. Diversity metric analysis

For 16S data, alpha diversity was assessed by three metrics in QIIME2: the observed number of OTUs (bacterial "richness"), Shannon's index (bacterial abundance and evenness of species present), and Faith's Phylogenetic Diversity index (Faith, 1992), which incorporates phylogenetic relatedness of taxa in each sample. Beta diversity was assessed using Bray-Curtis dissimilarity index, which measures abundance information, and the Jaccard similarity coefficient, which measures presence/absence information. Both metrics quantify the compositional dissimilarity between two different samples, bound between 0 and 1, where 0 is the same composition and 1 is maximally dissimilar composition. Metadata covariates were tested for associations with the microbiome using permutational multivariate analysis of variance (PERMANOVA) tests in R using the "adonis" function of the vegan package. PERMANOVA tests were done on both unweighted and weighted UniFrac distance matrices, which allows for comparison of intragroup and intergroup distances using a permutation scheme to obtain p-values.

PERMANOVA test were done with 999 randomizations. Low variance ASVs were removed for differential sample abundance analysis, which was determined with the edgeR (Robinson, McCarthy, & Smyth, 2010) and phyloseq (McMurdie & Holmes, 2013) packages in R. False discovery rate correction was performed on all resulting PERMANOVA and differential abundance p-values using the Benjamini-Hochberg (FDR/BH) criterion.

For shotgun metagenomic data, alpha diversity was calculated with Simpson's and Shannon's diversity indices. Simpson's diversity index is a measure of diversity which considers the number of species present, as well as the relative abundance of each species. The distribution of reads classified at the prokaryotic genus level and >1% abundance in each fecal sample ranged between a minimum of 195404 and a maximum of 7662130 reads. For diversity metrics, reads were randomly subsampled to 150000 reads. The R function vegdist in the vegan package (Okansen, 2013) computed dissimilarity indices using Bray-Curtis, which quantifies the compositional dissimilarity between two different samples.

4.6.8.4. Random forests

Random forest classifiers (RFCs), were implemented using the randomForest package (Liaw & Wiener, 2002) in R. Parameters included 5,001 decision trees, which were trained on taxa abundance data consisting of 14,138 ASVs for our 16S dataset and 20,844 taxa for our shotgun metagenomics dataset. Binary variables (e.g., positive, negative) were analyzed using classificatory RFC and continuous variables (e.g., cytokine values) were analyzed using regression RFC. Discriminating taxa were identified by random

forest using importance values, which were calculated as mean decrease in Gini index for classification random forests and percent increase in mean squared error (%IncMSE) for regression random forests. The top ten importance values are reported for each random forest test. Error of the model was assessed using out of bag (OOB) error. To increase the classifier's ability to detect true positives, we introduced a positive control consisting of statistical noise with a probability density equal to the values within the variable of interest (i.e., values within those present in the classification/regression variable) (https://github.com/marubel/R-ubelMisc). In RFC classification, prediction accuracy of the model was tested by randomly sampling half the samples of whichever of the two groups was smaller and training the classifier on this subset, for variables that had a minimum of at least five samples in each group. For example, if there were 80 positives and 100 negatives for a parasite, 40 positive and 40 negatives would be input as the training set for the classifier. RFC uses bootstrap sampling, which means that some training set samples in each downsampled category will be selected more than once, over a total of 5,001 iterations, to produce a consensus tree.

4.6.8.5. Co-occurrence analysis

Probabilistic co-occurrence analysis was done in R using the cooccur package (Griffith, Veech, & Marsh, 2016; Veech, 2013). Parasite pairs were removed if they shared less than one site. For parasite groups with presence/absence data, all pairwise combinations were tested using the hypergeometric distribution, which produced an observed-expected ratio and effect size for significant species combinations. Sample pairs were dubbed "random" if they didn't significantly differ from their expected number of co-occurrences

and if they did not deviate by <10% of the total number of sites, following the power analysis recommendations in Veech (2013). Pairwise combinations were visualized as heatmaps using ggplot2, where parasite combinations are measured from most negative to most positive interactions (left to right in the heatmap). Deviation from expected co-occurrence values was plotted against observed values.

4.6.8.6. Correlation analysis

Correlation analysis was conducted using the corrplot package in R (Wei & Simko, 2017). Correlations were calculated using Spearman's nonparametric rank-based correlation tests to control for potential outliers and hierarchical clustering was used to aggregate the correlation matrix. Correlation p- and r-values and figures are available in Supp. Tables 4-3, 4-8:4-10, and Supp. Figure 4-3 and 4-14. Correlation values were considered significant if p-values were less than or equal to 0.01. Cytokine plots incorporating 19 cytokines across 72 Cameroonians. For the metadata correlation analysis, we excluded 82 samples due to null values in metadata variables (n = 492 Cameroonians). Variables for the metadata correlation analysis were: Ants binary, Total Parasites, Total Parasites & HIV, Total Parasites & HIV & Blood Parasites, Total Parasites & Blood Parasites, Total ANTS Parasites, Height in centimeters, Weight in kilograms, Body Mass Index, Average Temperature, White blood cell count, Sex, Subsistence, Populations, Sampling site, HIV status, Neutrophile, Lymphocyte, Monocyte, Eosinophile, *Mf. M. perstans, P. falciparum, Mf. Loa loa, W. bancrofti, Microfilaria* spp., and Highly positive ANTS.

4.7. Declarations

4.7.1. Acknowledgements and funding

Many thanks to Boris Striepen, Robert Greenberg, Richard Marcantuno, Thomas J. Nolan, James Lok, and Ronald Collman for generously sharing parasite and pathogen samples for use as qPCR positive controls. Dan Beiting, members of the Penn CHoP microbiome core, the Penn Human Immunology Core, and the Bushman and Tishkoff labs gave invaluable project help and suggestions. We want to recognize Young Hwang (University of Pennsylvania), Jacob Leiby (University of Pennsylvania), Elizabeth Loy (University of Pennsylvania), Jaanki Dave (Geisinger Commonwealth School of Medicine), and Alexa Avitto (University of Pennsylvania) for assistance with extractions and assays. We thank André Essiane, Julius Y. Fonsah, Peter Kfu, Valentine N. Ndze, Eric N. Ngwang, and Grace N. Tenjei for field work in Cameroon; William Beggs and Lillian Chau for sample coordination; the Cameroonian villagers for their willingness to participate in this research; and the Cameroonian Ministry of Public Health and the Cameroon National Ethics Committee for permission to collect samples in Cameroon. We also thank community outreach groups, including the Mbororo Social and Cultural Association, for their support, as well as the partnership with the Faculty of Medicine and Biomedical Sciences of the University of Yaoundé 1, Cameroon. This research was supported in part by the Lewis and Clark Fund, the University of Pennsylvania, the Leakey Foundation, the Wenner-Gren Foundation (9299), an NIH training grant in Parasitology (5T32AI007532-18), and the National Science Foundation (BCS-1540432) to MAR and to SAT (BCS-1317217). Additional support came from NIH awards to SAT

(DP1 ES022577-04, 1R01DK104339-01 and 1R01GM113657-01) and to FB (R01-HL113252, R61-HL137063, U01-HL098957, R01-HL087115, K24-HL115354). FB also contributed assistance from the Penn Center for AIDS Research (P30-AI045008) and the PennCHoP Microbiome Program.

4.7.2. Availability of data and materials availability

16S amplicon and shotgun metagenomic sequence data are deposited under project accession PRJNA547591 in the NCBI Sequence Read Archive (SRA). Human Microbiome Project samples used for comparative analyses in shotgun data can be found in the SRA under the following accession numbers: SRR1804648, SRR1565914, SRR1803892, SRR1803862, SRR1804618, SRR1803903, SRR1803864, SRR1804203, SRR1803877, SRR1804107, SRR1804009, SRR1804055, SRR1804676, SRR532163, SRR1804148, SRR1804756, SRR1031154, SRR1804119, SRR1803355, SRR1803358, SRR1804539, SRR1564387, SRR512768, SRR1803287, SRR1031102, SRR1804688, SRR1804086.

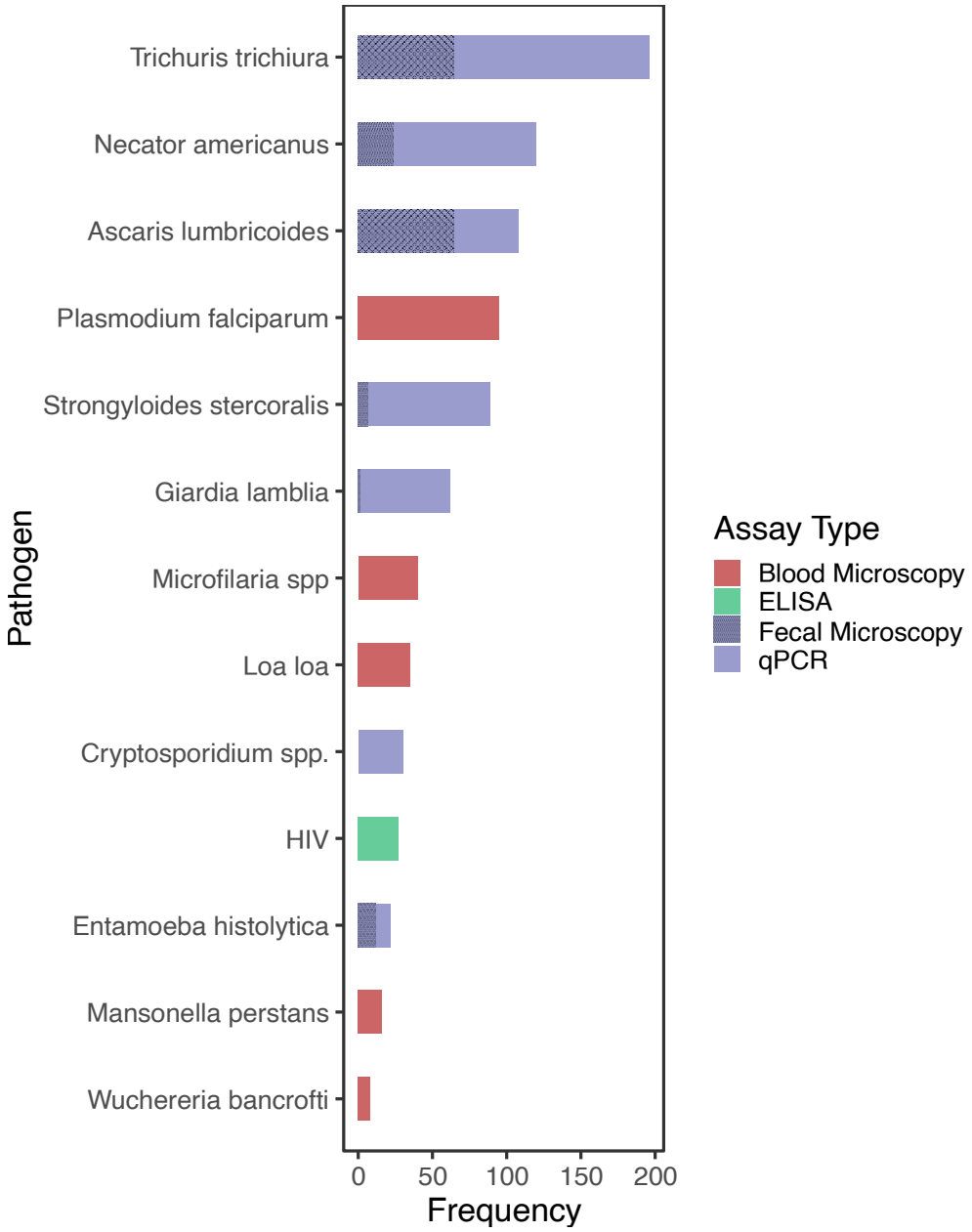
4.7.3. Authors' contributions

The study design was by SAT, FDB, and MAR. Cameroon biological samples and measurements were collected by EM, and MAR, with the coordination of SAT, CF, and AKN. U.S. fecal DNA (COMBO) were analyzed in collaboration with FDB. Fecal DNA extraction, 16S rRNA, shotgun sequencing, qPCRs, and HIV testing was conducted by MAR with support from FDB. MAR, LJT, AAA, and AC performed data analyses, with assistance from CT and KB. The manuscript was written by MAR with assistance from

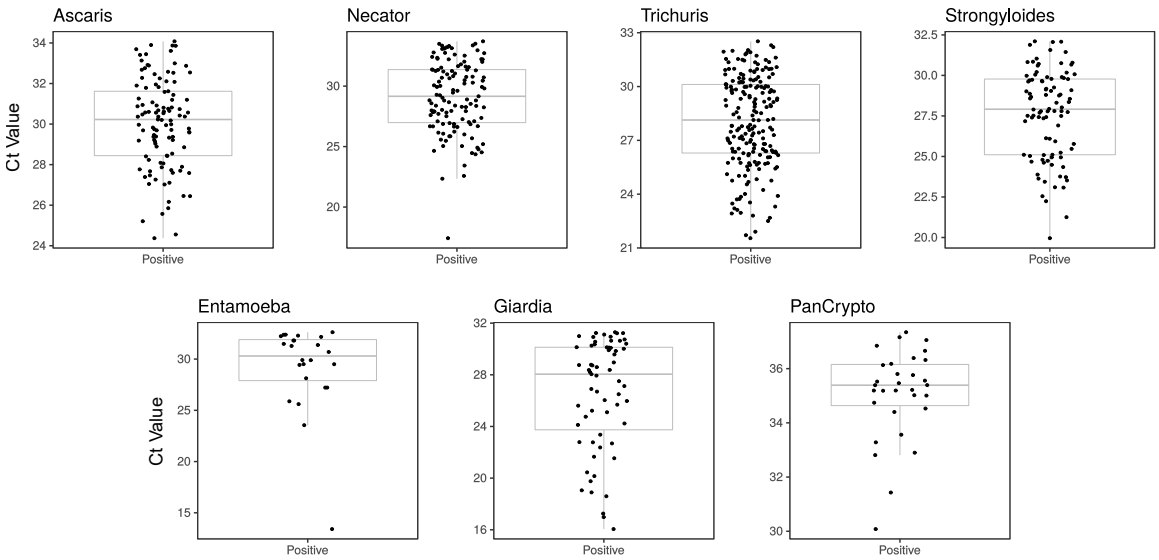
LJT and AAA, and was edited by SAT in consultation with FDB. All authors have read and approved the manuscript.

4.8. Supplemental Material

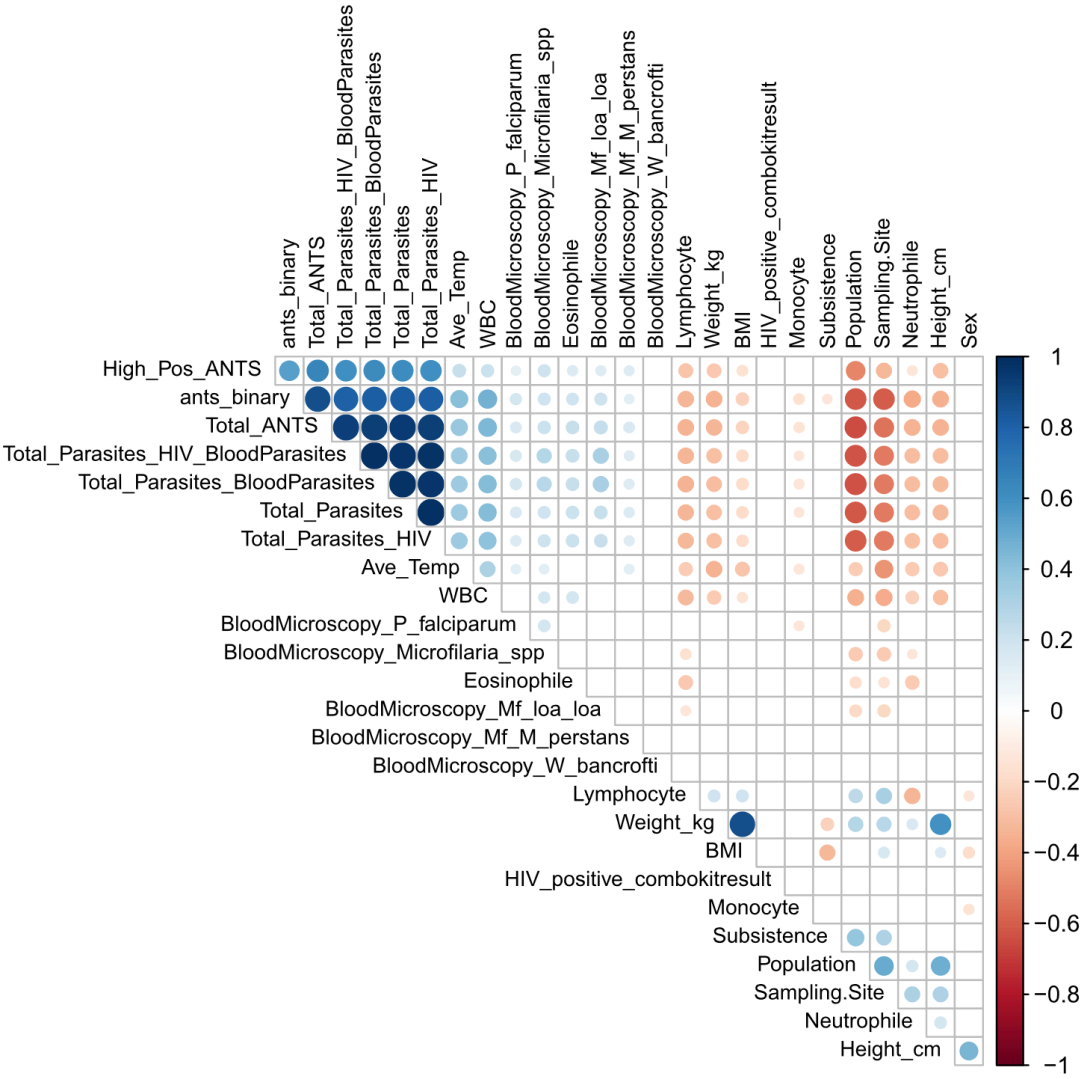
4.8.1. Supplemental Figures



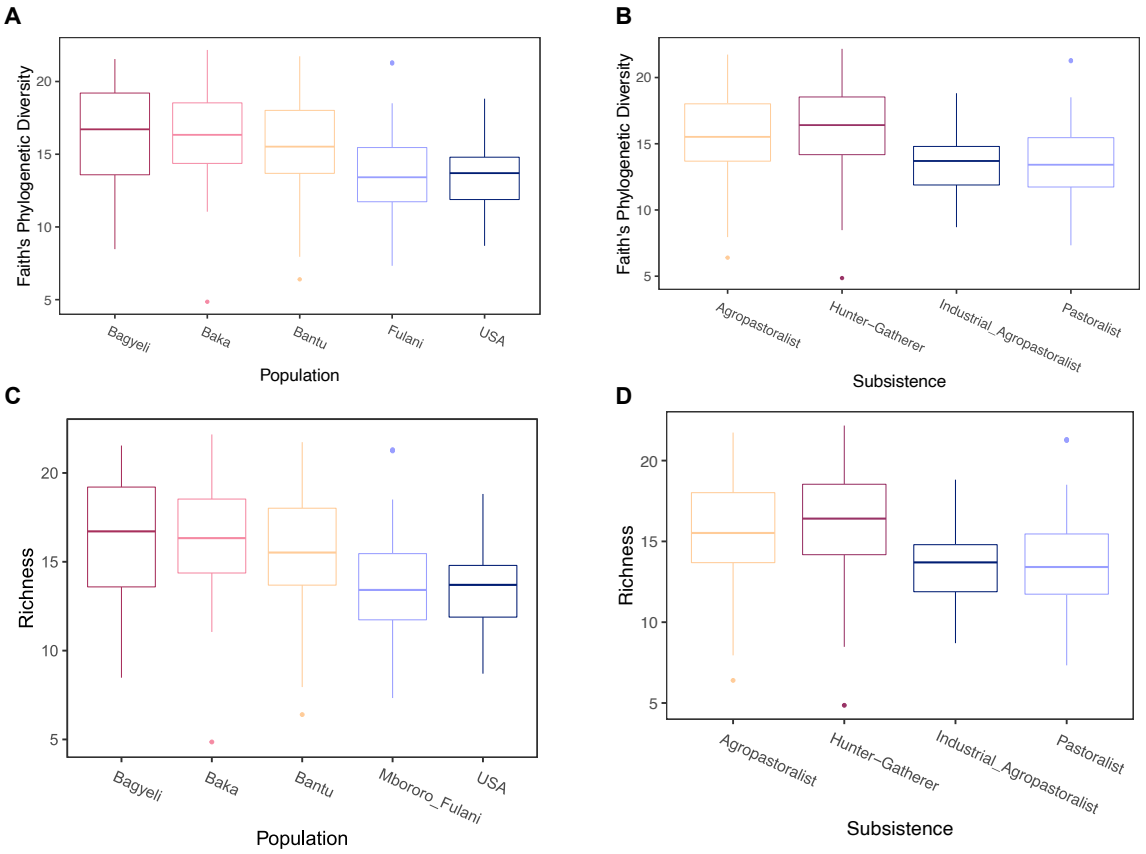
Supp. Figure 4-1. Infection frequency colored by assay type



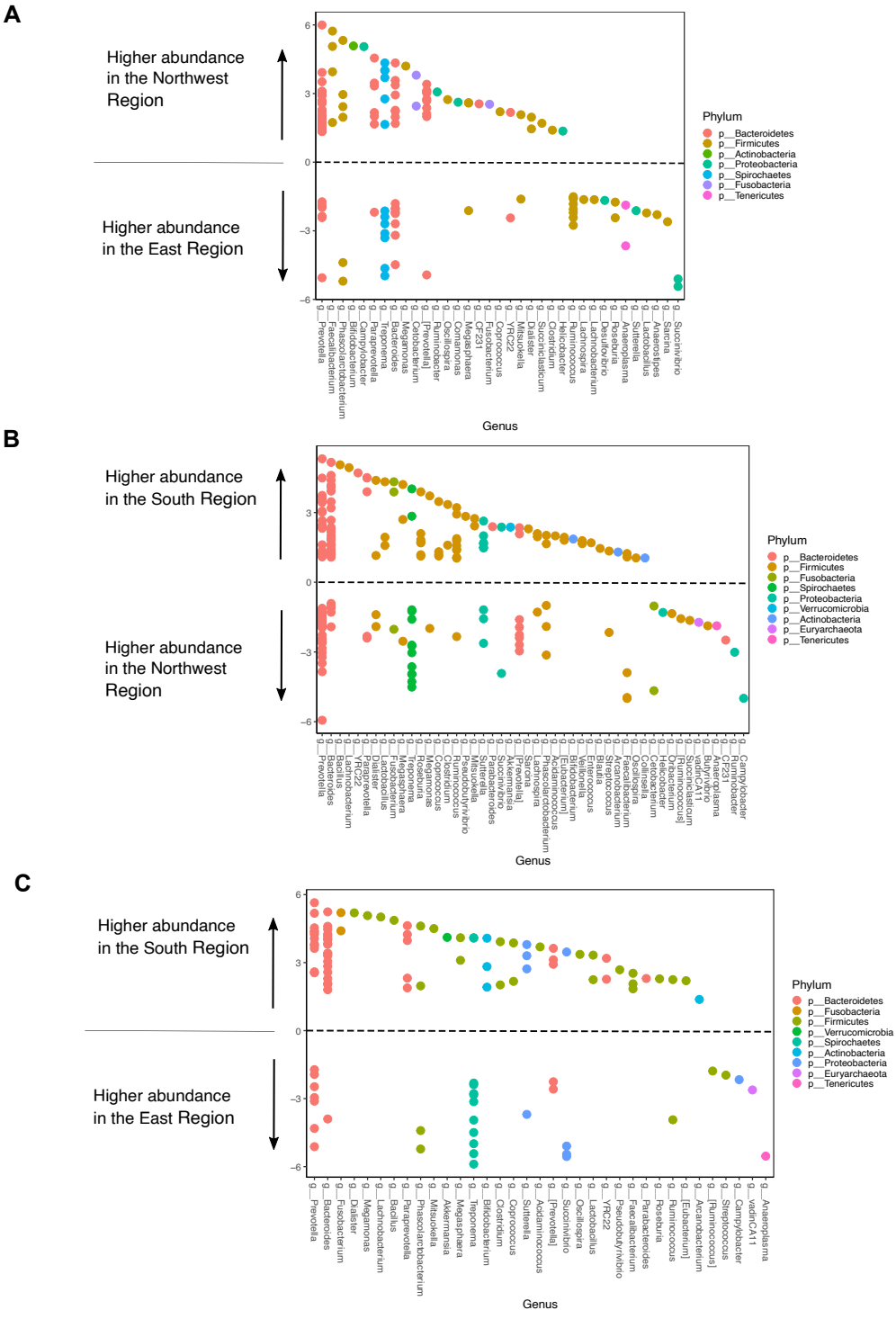
Supp. Figure 4-2. Positive parasite samples visualized by their qPCR cycle threshold values (Ct), with ANTS parasites on the top row



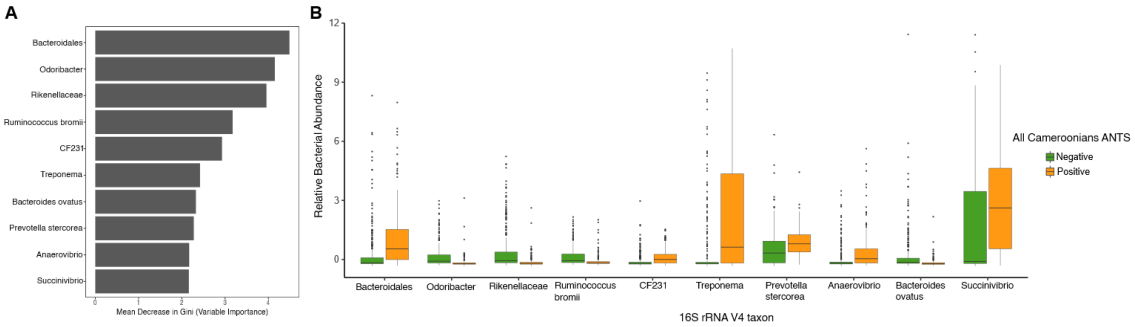
Supp. Figure 4-3. Correlation plot between metadata variables of interest



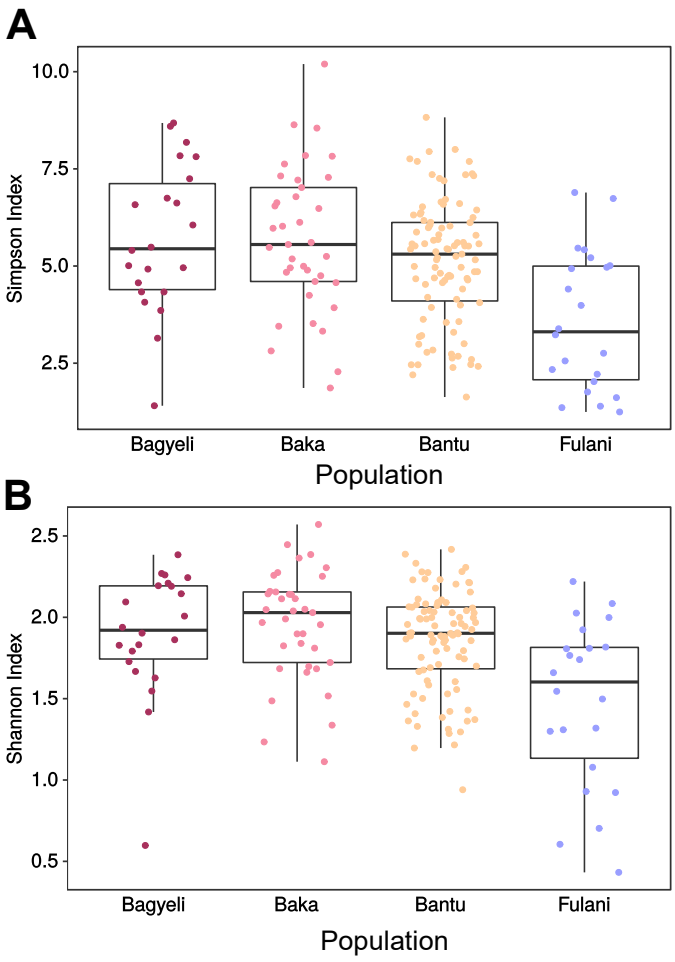
Supp. Figure 4-4. Bacterial 16S rRNA alpha diversity metrics across populations and subsistence groups. A) Faith's Phylogenetic Diversity for U.S. and Cameroon populations. B) Faith's Phylogenetic Diversity for subsistence groups. C) Richness for U.S. and Cameroon populations. D) Richness for subsistence groups.



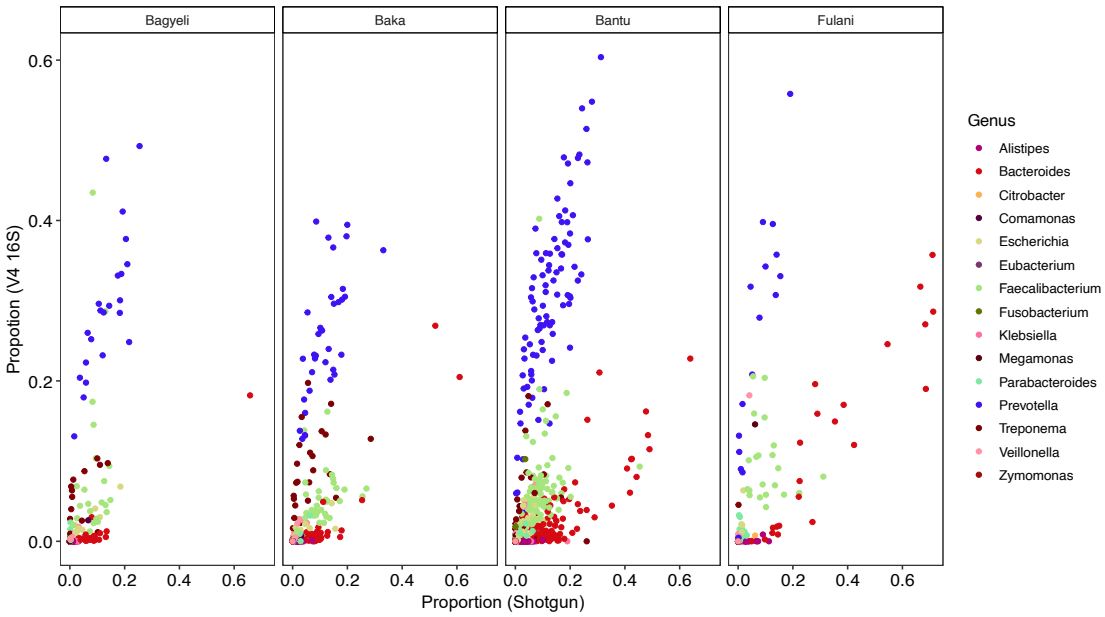
Supp. Figure 4-5. Most differentially abundant taxa between Bantu regions by bacterial phylum for A) Bantu in Northwest and East regions, B) Bantu in the South and Northwest Regions, and C) Bantu in the South and East Regions.



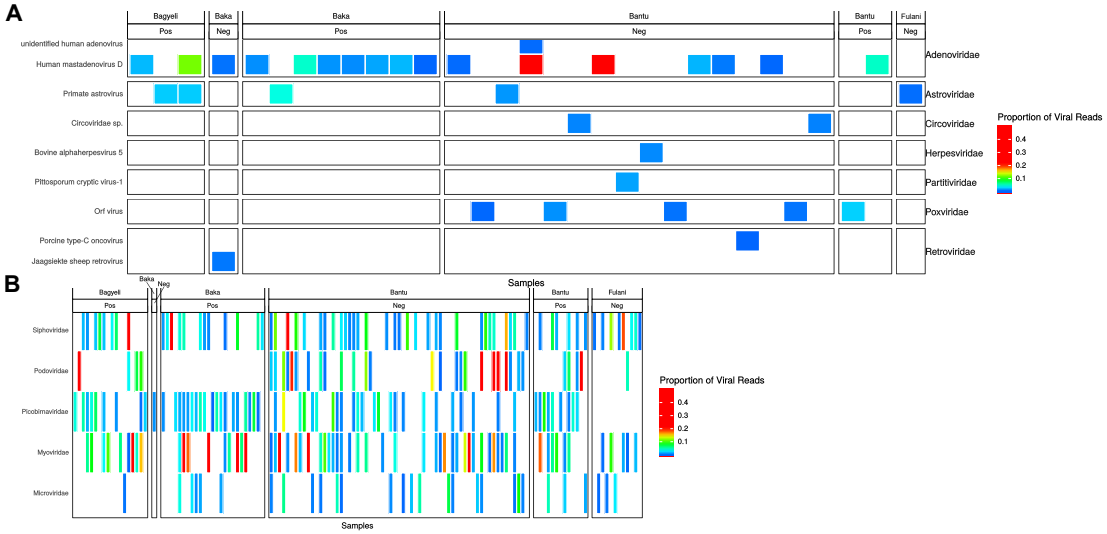
Supp. Figure 4-6. Top ten most important variables in RFC from 16S rRNA V4 analysis for all Cameroonians with and without ANTS parasites (left), and relative abundances of RFC taxa visualized between positive and negative Cameroonians (right)



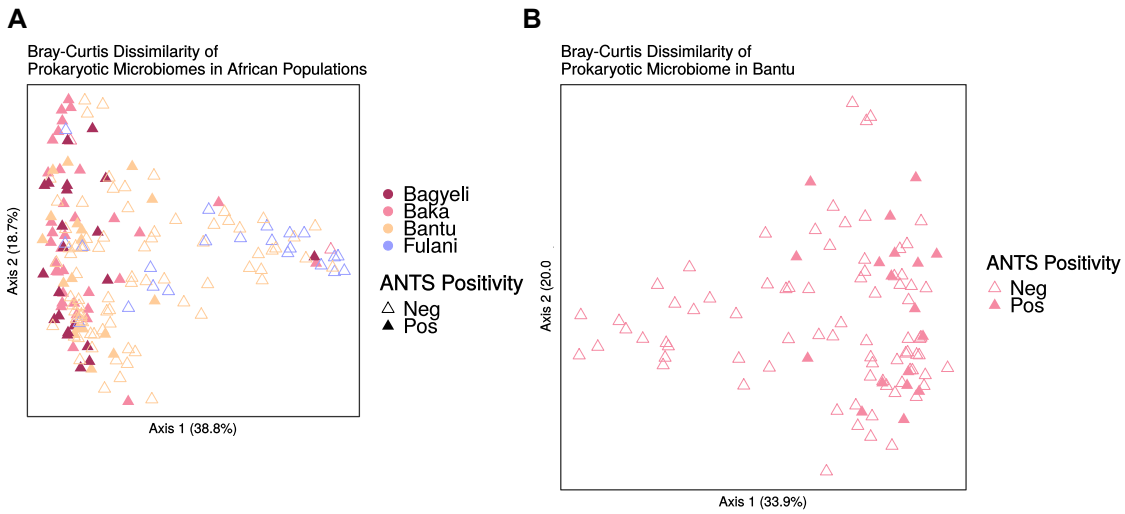
Supp. Figure 4-7. Shotgun sequencing diversity metrics across Cameroonian populations. A) Simpson Index across all populations and B) Shannon index across all populations.



Supp. Figure 4-8. Comparison of the most abundant 15 taxa proportions in shotgun sequencing and 16S rRNA amplicon sequencing within each Cameroonian population

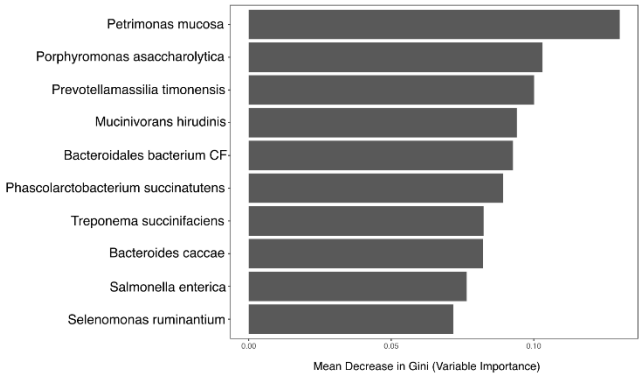


Supp. Figure 4-9. Viral annotation from shotgun metagenomic data
A) Taxonomy and proportion of viral reads in ANTS positive and negative subgroupings of each Cameroonian population. B) Taxonomy and proportion of bacteriophage reads in ANTS positive and negative subgroupings of each Cameroonian population.

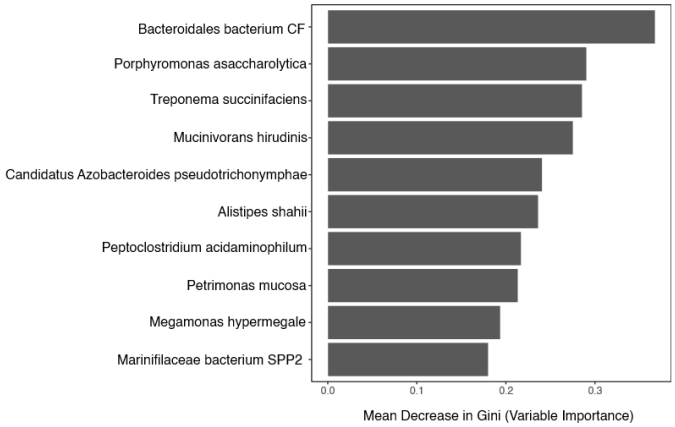


Supp. Figure 4-10. Bray-Curtis dissimilarity of prokaryotic microbiomes shaded by ANTS positive status for A) All Cameroonian populations and B) in Bantu only.

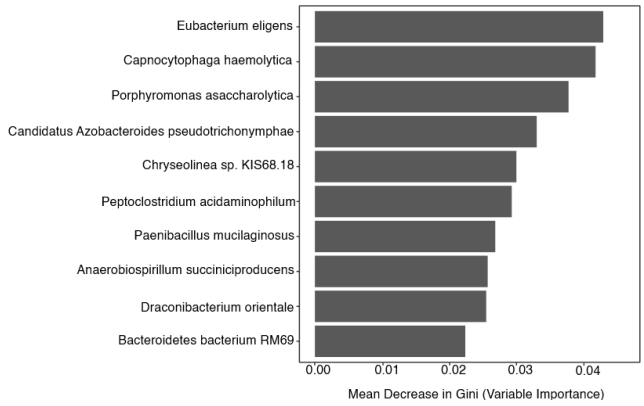
A



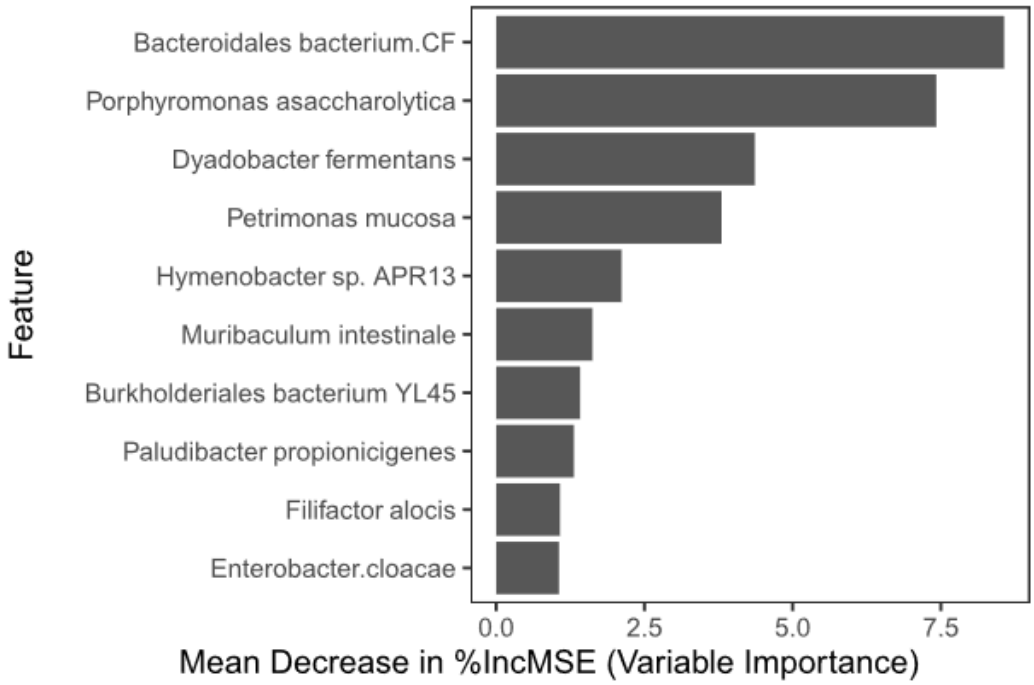
B



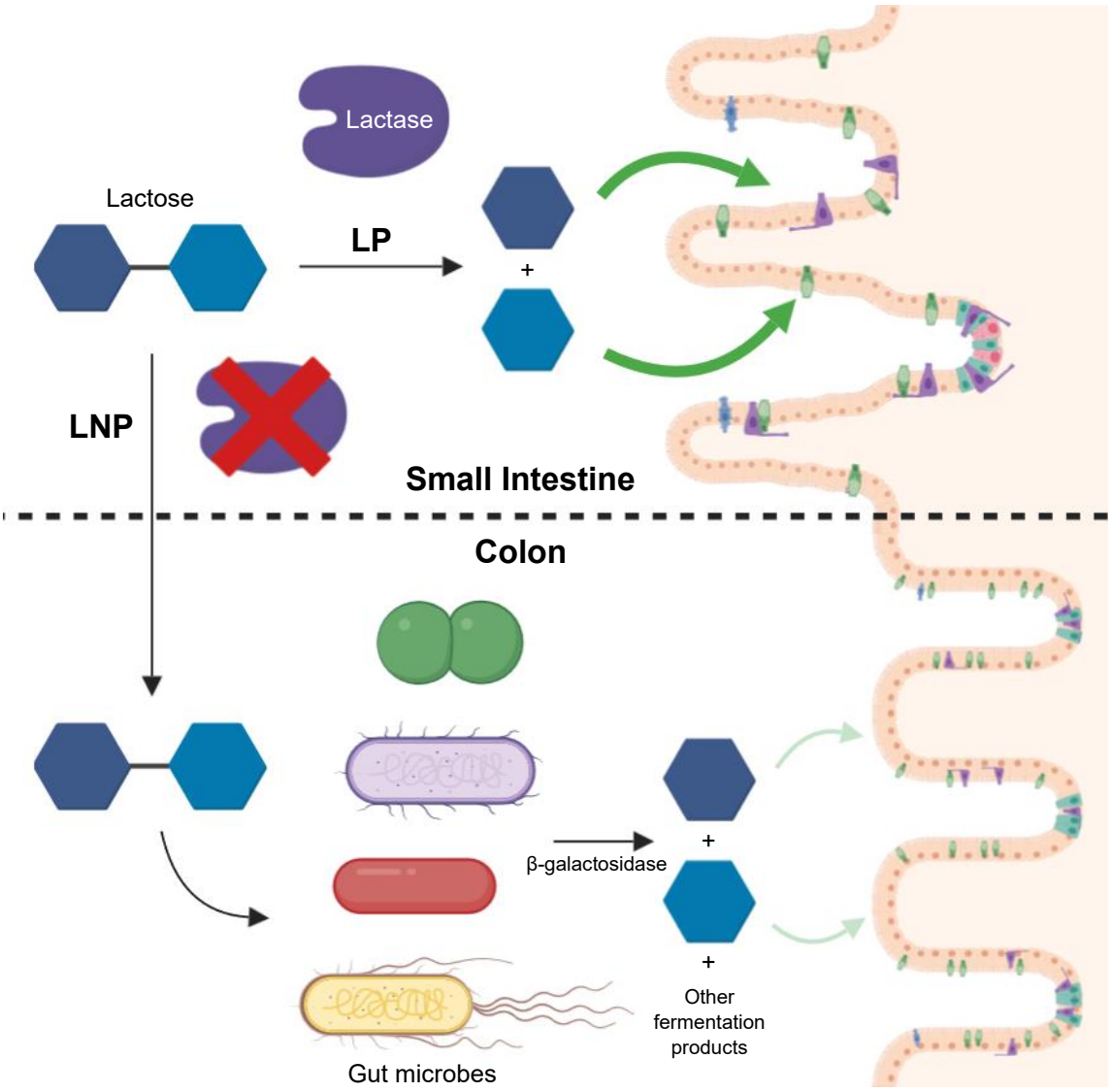
C



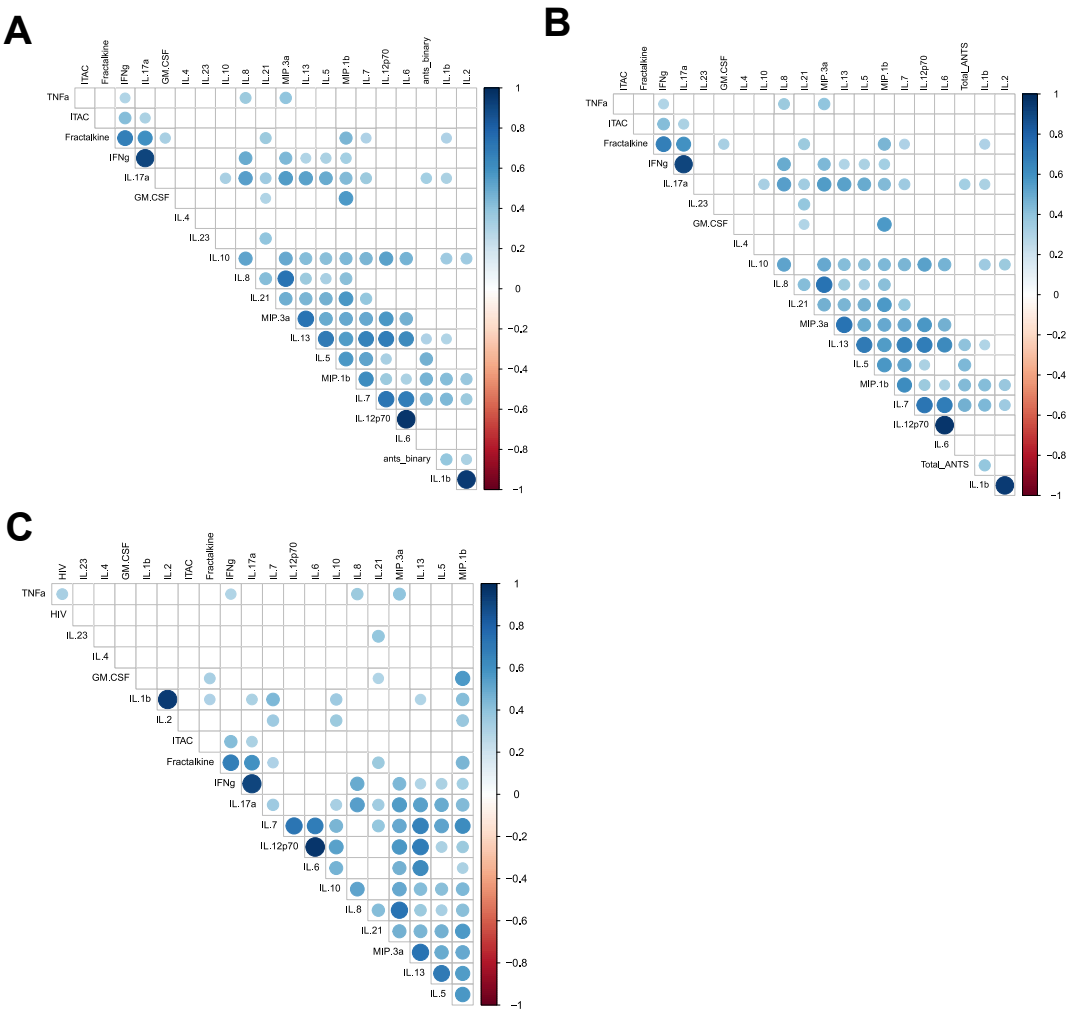
Supplemental Figure 11. Top ten most significant taxa in RFCs on shotgun data for A) Subsistence for all Cameroonians B) ANTS infection (positive/negative) in all Cameroonians and C) ANTS infection (positive/negative) in Bantu only.



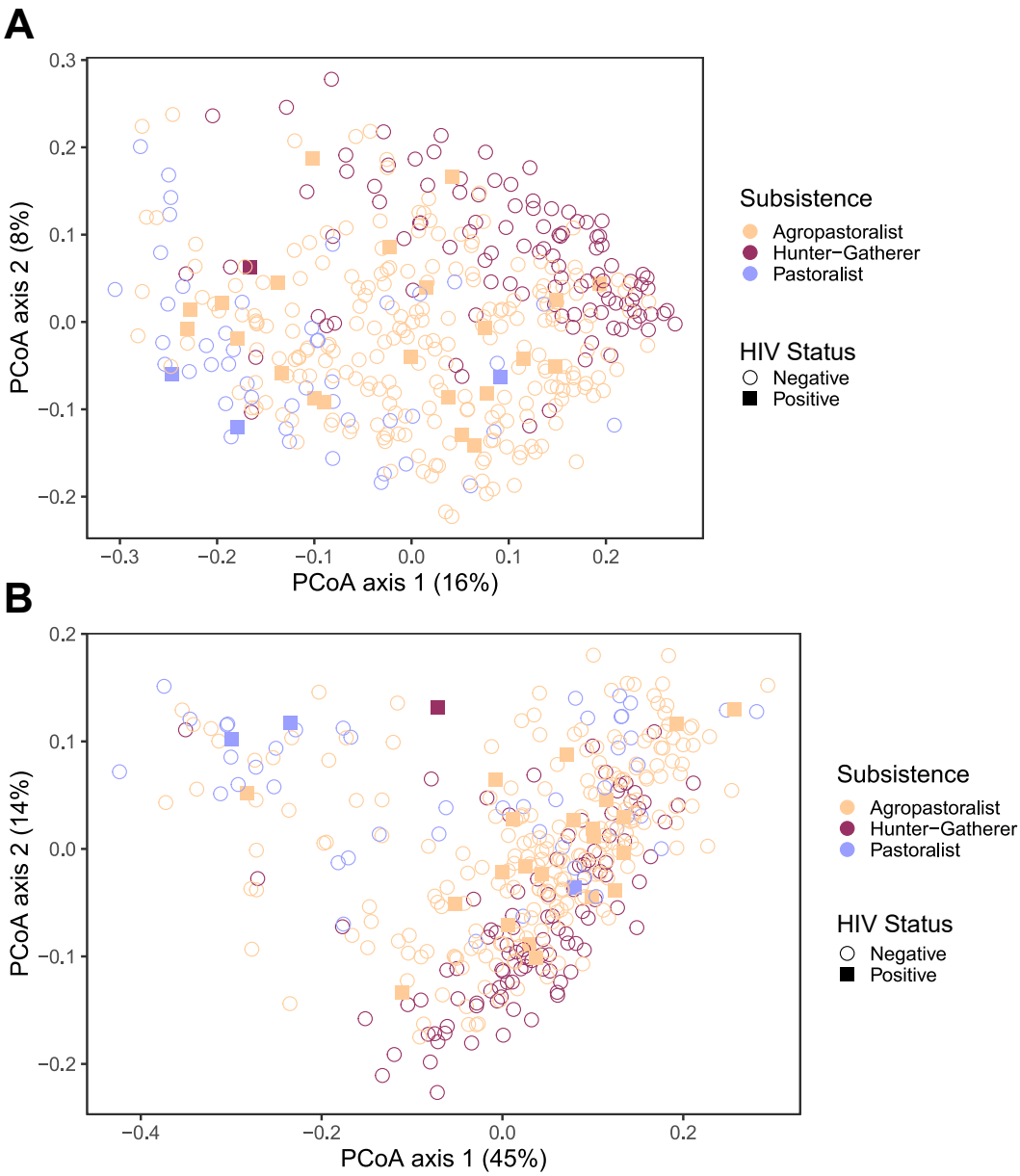
Supp. Figure 4-12. Top ten most significant taxa in RFCs on shotgun data for IL-5



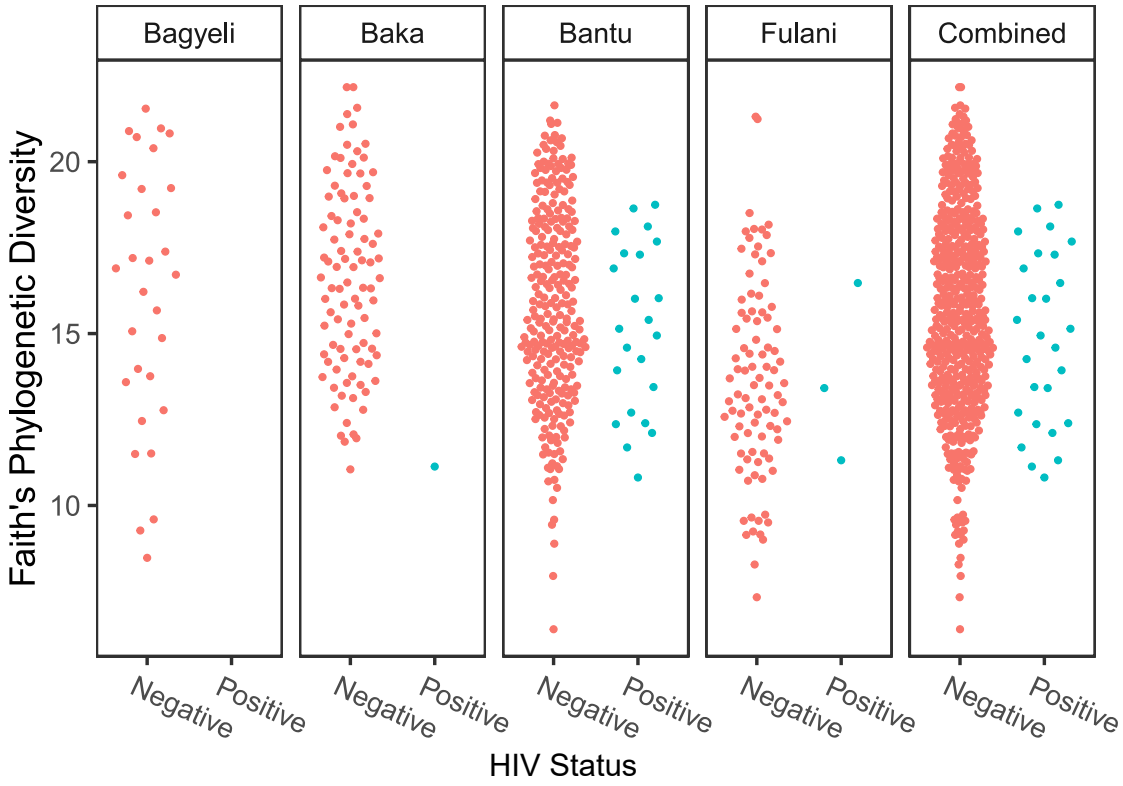
Supp. Figure 4-13. Lactose metabolism pathways contingent on host lactase phenotype. If humans produce the lactase enzyme after weaning, they can metabolize lactose sugar from dairy. This phenotype is called lactase persistence (LP). In LP individuals (top of the dashed line), the disaccharide lactose will be cleaved by the lactase enzyme into monosaccharides glucose and galactose at the brush-border of the small intestine. Glucose and galactose are transferred into absorptive enterocytes, where they then move through the intestinal lumen, across the epithelium and enter the host bloodstream for further metabolism. If humans cease producing the lactase enzyme after weaning, they are not capable of breaking down lactose sugar in the small intestine. This phenotype is called lactase non-persistence (LNP) (bottom of dashed line). In this case, ingested lactose will pass through the small intestine and enter the large intestine. Bacteria who possess the LAC operon are capable of producing the enzyme β-galactosidase. Host colonic bacteria producing β-galactosidases are also capable of catabolizing lactose into glucose and galactose. The fermentation products of this reaction can produce a series of intermediate (e.g., succinate, lactate, and formate) and end-product metabolites (butyrate, propionate, acetate, gases CO_2, H_2, and CH_4).



Supp. Figure 4-14. Correlation plots between cytokine values and pathogen variables of interest. Negative correlations are displayed in red and positive correlations are shown in blue. Color intensity and the circle size are proportional to correlation coefficients. A) Correlation plot of ANTS infection ("ants_binary") with cytokines. B) Correlation plot of ANTS counts ("Total_ANTS") with cytokines. C) Correlation plot of HIV status ("HIV") with cytokines. D: Top ten most significant taxa in regression RFCs on shotgun data for IL-5 cytokine values.



Supp. Figure 4-15. Principal Coordinates Analysis (PCoA) on A) unweighted and B) weighted bacterial UniFrac distances, shaded by HIV status. Data points colored by Cameroon subsistence groups, with filled squares denoting HIV positive individuals, for A: Unweighted, or bacterial presence/absence, PCoA and B: Weighted, or bacterial abundance, PCoA



Supp. Figure 4-16. Bacterial alpha diversity metrics across populations and HIV status, with dots representing individual samples

4.8.2. Supplemental tables and files

Supplemental Tables:

All supplemental tables are attached in the Digital Supplement named

SupplementalTables_Ch4.xlsx.

Supp. Table 4-1. Metadata.

Supp. Table 4-2. Co-occurrence statistical analysis.

Supp. Table 4-3. Correlation plots with correlation coefficients and Spearman's nonparametric rank-based correlation test p-values for metadata variables.

Supp. Table 4-4. Unweighted Adonis (PERMANOVA) statistical tests on 16S bacterial data.

Supp. Table 4-5. Weighted Adonis (PERMANOVA) statistical tests on 16S bacterial data.

Supp. Table 4-6. Bacterial 16S rRNA and shotgun sequencing classification and regression RFCs on metadata variables with model accuracy, individual taxa values in model importance, and taxa names.

Supp. Table 4-7. Pathway maps from shotgun data for taxa identified in RFC analysis on ANTS status.

Supp. Table 4-8. Correlation plots with correlation coefficients and Spearman's nonparametric rank-based correlation test p-values for ANTS infection (positive or negative, "ANTS binary") and cytokine values

Supp. Table 4-9. Correlation plots with correlation coefficients and Spearman's nonparametric rank-based correlation test p-values for total ANTS (counts of ANTS parasites) and cytokine values

Supp. Table 4-10. Correlation plots with correlation coefficients and Spearman's nonparametric rank-based correlation test p-values for HIV status (positive or negative) and cytokine values.

Supp. Table 4-11. Odds ratios.

Supp. Table 4-12. Organismal source and details of V4 positive control synthetic DNA sequences.

Supp. Table 4-13. Details on primers and probes used to target different parasites in qPCR.

Supp. Table 4-14. Metrics from the HiSeq machine following the shotgun sequencing run.

Supp. Table 4-15. Quality scores on fastq files (forward and reverse reads) across nucleotide reads for samples sequenced in shotgun analysis.

Supp. Table 4-16. Reads per sample across different phases of preprocessing, including those that matched to the human reference sequence, which were removed from analysis.

5. Conclusions and Future Directions

In this thesis, I characterized gut microbiome variation across diverse, rural African populations in three countries and spanning multiple subsistence practices. Chapter 2 provided a background on the complex factors associated with gut microbiome composition across global populations, and then contextualized this background within the evolutionary history and contemporary health of traditional sub-Saharan Africans. Chapters 3 and 4 tested a multitude of these factors, including pathogens, host genetics and physiology, and geography, for association with the human gut microbiome. Importantly, the same experimental and analytic pipeline was used in each study across multiple populations, and the full spectrum of microbiota was interrogated in Cameroonians.

In Chapter 3, 16S amplicon sequencing showed that African and U.S. gut microbiomes primarily differed in *Bacteroidaceae* and *Prevotellaceae* abundances and their higher overall diversity. However, there was a group of Botswana Bantu who were *Bacteroidaceae* high, even after controlling for age, sex, and sampling site. Overall, the Bantu had higher BMI than other populations tested in Tanzania and Botswana, but BMI did not explain the differences in the *Bacteroides*-high group. Botswanans had an enrichment of bacterial genes implicated in the degradation of xenobiotics and industrial pollutants. I speculated that the differences seen in these Bantu may be reflective of subsistence transitions and "soft" measures of industrialization that remain to be measured, such as increased proximity through travel to urban areas and increased exposure to chemicals and clinical care.

I found sex specific differences and a high amount of diversity in hunter-gatherer microbiomes, who possessed high amounts of fiber-degrading *Treponema* bacteria. Surprisingly, there were no significant differences between pastoralists and agropastoralists gut microbiomes, as had been originally hypothesized. Instead, some of the starkest differences between Botswanan and Tanzanian microbiomes stemmed from geography- that is, Botswanans and Tanzanians had gut microbiomes that were more similar to populations within their own country than populations outside their country, regardless of subsistence. Concordantly, the presence or absence of bacterial taxa was strongly associated with geography- meaning that Botswanans and Tanzanians tended to be colonized by the microbiota of their local geography. However, the abundance of a bacterial taxon was strongly associated with both geography and host genetics, which indicated that once established in the gut, the abundance of a taxon may be a heritable trait.

In Chapter 4, I paired metagenomic sequencing, amplicon sequencing, a multitude of tests for pathogens, and anthropometrics to interrogate the gut microbiomes of Cameroonian hunter-gatherers, agropastoralists, and pastoralists with comparative populations from the U.S. Although blood parasitemia and HIV were considered in initial analyses, neither had as significant an impact on gut microbiome composition as ANTS parasite infection. After the gut microbiome differences associated with Cameroon versus the U.S. (again being a difference in *Prevotellaceae* and *Bacteroidaceae* abundance), one of the most statistically robust correlation between the microbiome and any of our tested variables was with gastroenteric parasites. In particular, a significantly co-occurring suite

of soil-transmitted helminths, the “ANTS” group, could be predicted with high accuracy using gut microbiome composition from both amplicon and shotgun sequences using supervised machine-learning methods. Th2 and Th1 responses varied among the sampled individuals, but the cytokine IL-5 was positively correlated with ANTS infection. RFC models implicated overlapping taxa as being highly predictive of both IL-5 levels and multiple ANTS categories. Whether or not these associated microbial taxa were present prior to infection, were introduced by the parasite, or were induced because of the parasite (from its own microbiota), remains an open question. The Cameroonian populations were sampled in the dry season through the beginning of the rainy season, when parasite rates increase. Thus, the infection rates and microbial correlations discussing in this research may underrepresent the maximum annual incidence and the extent of their co-association. . Longitudinally sampling the same individuals across multiple timepoints will help address these questions and will substantiate the effects of long-term inter-individual variation within the gut microbiomes of parasitized individuals.

Future directions

This research centered around amplicon and shotgun sequencing of human gut microbiota. Amplicon sequencing is both more cost effective and scalable; however, it affords little in the way of fully analyzing and categorizing genome function without having to resort to imputation methods, and has no way to discern lateral gene transfer or mutation within an amplicon sequence. Although more costly, whole genome shotgun sequencing methods allow the sequencing of partial to full genomes across domains of

life within a sample, and as such, can be useful for sequencing and annotating whole microbial genomes. As the cost of shotgun sequencing continues to decrease, the volume of studies favoring this technology instead of amplicon sequencing is bound to rise. Increasingly, shotgun metagenomics methods are paired with other '-omics technologies, including host metabolomics, microbial metabolomics (sometimes called meta-metabolomics), and RNA-seq to develop an integrated understanding of total organismal interactions. One promising area is the use of techniques that amplify virus like particles within fecal DNA to yield greater viral detection on sequencing runs. Limited virome work has been done in Cameroon and none, as of this writing, has specifically focused on the human gut virome in response to industrialization and subsistence changes, or the virome's associations with gastroenteric parasites. Such integrated studies have, and will likely continue, to change our understanding of the essential role that microbes play in the biology of humans.

Furthermore, the application of novel methods to classify unknown sequences, of which we had many in our African populations, will expand the pangenome of contemporary human-associated commensal and pathogenic gut microbes (Pasolli et al., 2019). Improvement in next-generation sequencing and computational methods may yield insights into ancient human microbiota from even older timescales than can presently be queried, using diverse samples such as fossilized coprolites, preserved tissue (Maixner et al., 2016), and mineralized microbial calculus from teeth (Warinner et al., 2014). Integrating ancient microbiome sequencing with that of contemporary groups would allow the testing of hypotheses about gut microbiome evolutionary changes, including

timing of industrialization changes over human history, and the extent that such changes are globally ubiquitous.

These integrated research studies have not yet described the endogenous gastroenteric human parasite microbiome, or investigated to what extent parasite microbiota may contribute to pathogenicity or immune response in the host. It's unknown whether gastroenteric parasites themselves directly stimulate host immune responses or indirectly signal host immune responses through gut microbiota. If the latter, antihelminthic/anti-parasitic treatments could be developed that target the gut microbiome as an intermediate step in the interruption of the parasite life cycle. Recently, the "Parasite Microbiome Project" was proposed to research the role of microbiota in host-parasite dynamics (Dheilly et al., 2017). Hopefully, projects such as these will act as a clarion call for research groups to develop comparable analyses across model and human systems, and for national agencies and foundations to prioritize the funding of these proposals.

There is some evidence to indicate that traditional populations immigrating to the U.S. lose their microbial diversity almost immediately, and that this reduction in diversity is associated with a rise in obesity (Vangay et al., 2018). This microbial diversity, once lost, may not be regained in successive generations (Sonnenburg et al., 2016). Although overlooked in many studies of microbial diversity, the loss of both pathogenic and commensal eukaryotes has uncertain implications for human health. We found that gut microbiome composition could be used to predict commensal and pathogenic *Entamoeba* together with more accuracy than either group separately. Eukaryotes such as *Blastocystis* remain relatively unexplored in their relationship to human microbiota.

Finally, efforts to preserve human microbial diversity from global populations through biobanking initiatives are likely to gain widespread traction in the coming decade. These already exist within universities and some private companies, and may soon include traditional populations (Bello et al., 2018). Infrastructure building in industrializing countries, obtaining informed consent, and the returning of results to research participants are just the beginning of the responsibility scientists have to their research participants. The issue of who may own the commercial products made from traditional microbiome strains is ethically fraught, particularly since many traditional populations are marginalized, resource-poor, and may not fully understand the long-term implications of contributing a biobank specimen to a commercial enterprise. However, when traditional groups are studied, it is of critical importance to bring the voices of these communities into collaboration as active, informed participants in the research process, so that responsible research outcomes can be generated.

BIBLIOGRAPHY

Aas, J., Gessert, C. E., & Bakken, J. S. (2003). Recurrent Clostridium difficile Colitis: Case Series Involving 18 Patients Treated with Donor Stool Administered via a Nasogastric Tube. *Clinical Infectious Diseases*, *36*(5), 580–585. https://doi.org/10.1086/367657

Abbas, A. A., Taylor, L. J., Dothard, M. I., Leiby, J. S., Fitzgerald, A. S., Khatib, L. A., … Bushman, F. D. (2019). Redondoviridae, a Family of Small, Circular DNA Viruses of the Human Oro-Respiratory Tract Associated with Periodontitis and Critical Illness. *Cell Host & Microbe*, *25*(5), 719-729.e4. https://doi.org/10.1016/j.chom.2019.04.001

Adjemian, J., Farnon, E. C., Tschioko, F., Wamala, J. F., Byaruhanga, E., Bwire, G. S., … Rollin, P. E. (2011). Outbreak of Marburg hemorrhagic fever among miners in Kamwenge and Ibanda Districts, Uganda, 2007. *The Journal of Infectious Diseases*, *204 Suppl 3*, S796-799. https://doi.org/10.1093/infdis/jir312

Africa, W. R. O. for. (2006). *Health of the People The African Regional Health Report* (1 edition). Brazzaville, Republic of Congo: World Health Organization.

Akatsu, H., Iwabuchi, N., Xiao, J.-Z., Matsuyama, Z., Kurihara, R., Okuda, K., … Maruyama, M. (2013). Clinical effects of probiotic Bifidobacterium longum BB536 on immune function and intestinal microbiota in elderly patients receiving enteral tube feeding. *JPEN. Journal of Parenteral and Enteral Nutrition*, *37*(5), 631–640. https://doi.org/10.1177/0148607112467819

Alkorta-Aranburu, G., Beall, C. M., Witonsky, D. B., Gebremedhin, A., Pritchard, J. K., & Di Rienzo, A. (2012). The Genetic Architecture of Adaptations to High Altitude in Ethiopia. *PLoS Genetics*, *8*(12). https://doi.org/10.1371/journal.pgen.1003110

Allen, J. E., & Maizels, R. M. (2011). Diversity and dialogue in immunity to helminths. *Nature Reviews. Immunology*, *11*(6), 375–388. https://doi.org/10.1038/nri2992

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, *59*(1), 143–169.

Anderson, M. (2001). A new method for non-parametric multivariate analysis of variance. *Australian Ecology*, *26*, 32–46.

Angelakis, E., Bachar, D., Yasir, M., Musso, D., Djossou, F., Gaborit, B., … Raoult, D. (2018). Treponema species enrich the gut microbiota of traditional rural populations but are absent from urban individuals. *New Microbes and New Infections*, *27*, 14–21. https://doi.org/10.1016/j.nmni.2018.10.009

Annunziato, F., Romagnani, C., & Romagnani, S. (2015). The 3 major types of innate and adaptive cell-mediated effector immunity. *The Journal of Allergy and Clinical Immunology*, *135*(3), 626–635. https://doi.org/10.1016/j.jaci.2014.11.001

Ardissone, A. N., de la Cruz, D. M., Davis-Richardson, A. G., Rechcigl, K. T., Li, N., Drew, J. C., … Neu, J. (2014). Meconium microbiome analysis identifies bacteria correlated with premature birth. *PloS One*, *9*(3), e90784. https://doi.org/10.1371/journal.pone.0090784

Århem, K. (1989). The Cultural Connotations of Milk, Meat, and Blood in the Pastoral Maasai Diet. *Anthropos*, (84), 1–23.

Arola, H. (1994). Diagnosis of hypolactasia and lactose malabsorption. *Scandinavian Journal of Gastroenterology. Supplement*, *202*, 26–35.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., … Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, *473*(7346), 174–180. https://doi.org/10.1038/nature09944

Aubert, M., Pike, A. W. G., Stringer, C., Bartsiokas, A., Kinsley, L., Eggins, S., … Grün, R. (2012). Confirmation of a late middle Pleistocene age for the Omo Kibish 1 cranium by direct uranium-series dating. *Journal of Human Evolution*, *63*(5), 704–710. https://doi.org/10.1016/j.jhevol.2012.07.006

Audebert, C., Even, G., Cian, A., Blastocystis Investigation Group, Loywick, A., Merlin, S., … Chabé, M. (2016). Colonization with the enteric protozoa Blastocystis is associated with increased diversity of human gut bacterial microbiota. *Scientific Reports*, *6*, 25255. https://doi.org/10.1038/srep25255

Ayeni, F. A., Biagi, E., Rampelli, S., Fiori, J., Soverini, M., Audu, H. J., … Turroni, S. (2018). Infant and Adult Gut Microbiome and Metabolome in Rural Bassa and Urban Settlers from Nigeria. *Cell Reports*, *23*(10), 3056–3067. https://doi.org/10.1016/j.celrep.2018.05.018

Azad, M. B., Konya, T., Maughan, H., Guttman, D. S., Field, C. J., Sears, M. R., … Kozyrskyj, A. L. (2013). Infant gut microbiota and the hygiene hypothesis of allergic disease: Impact of household pets and siblings on microbiota composition and diversity. *Allergy, Asthma, and Clinical Immunology: Official Journal of the*

Canadian Society of Allergy and Clinical Immunology, *9*(1), 15. https://doi.org/10.1186/1710-1492-9-15

Azcarate-Peril, M. A., Ritter, A. J., Savaiano, D., Monteagudo-Mera, A., Anderson, C., Magness, S. T., & Klaenhammer, T. R. (2017). Impact of short-chain galactooligosaccharides on the gut microbiome of lactose-intolerant individuals. *Proceedings of the National Academy of Sciences*, *114*(3), E367–E375. https://doi.org/10.1073/pnas.1606722113

Bache, S. M., & Wickham, H. (2014). *magrittr: A Forward-Pipe Operator for R*. Retrieved from https://CRAN.R-project.org/package=magrittr

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., … Wang, J. (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host & Microbe*, *17*(6), 852. https://doi.org/10.1016/j.chom.2015.05.012

Bak, S. H., Choi, H. H., Lee, J., Kim, M. H., Lee, Y. H., Kim, J. S., & Cho, Y.-S. (2017). Fecal microbiota transplantation for refractory Crohn's disease. *Intestinal Research*, *15*(2), 244–248. https://doi.org/10.5217/ir.2017.15.2.244

Bakken, J. S., Borody, T., Brandt, L. J., Brill, J. V., Demarco, D. C., Franzos, M. A., … Surawicz, C. (2011). Treating Clostridium difficile Infection With Fecal Microbiota Transplantation. *Clinical Gastroenterology and Hepatology*, *9*(12), 1044–1049. https://doi.org/10.1016/j.cgh.2011.08.014

Bancroft, A. J., Hayes, K. S., & Grencis, R. K. (2012). Life on the edge: The balance between macrofauna, microflora and host immunity. *Trends in Parasitology*, *28*(3), 93–98. https://doi.org/10.1016/j.pt.2011.12.001

Bär, A.-K., Phukan, N., Pinheiro, J., & Simoes-Barbosa, A. (2015). The Interplay of Host Microbiota and Parasitic Protozoans at Mucosal Interfaces: Implications for the Outcomes of Infections and Diseases. *PLoS Neglected Tropical Diseases*, *9*(12). https://doi.org/10.1371/journal.pntd.0004176

Barroso, E., Van de Wiele, T., Jiménez-Girón, A., Muñoz-González, I., Martín-Alvarez, P. J., Moreno-Arribas, M. V., … Requena, T. (2014). Lactobacillus plantarum IFPL935 impacts colonic metabolism in a simulator of the human gut microbiota during feeding with red wine polyphenols. *Applied Microbiology and Biotechnology*, *98*(15), 6805–6815. https://doi.org/10.1007/s00253-014-5744-1

Bartram, H. P., Scheppach, W., Gerlach, S., Ruckdeschel, G., Kelber, E., & Kasper, H. (1994). Does yogurt enriched with Bifidobacterium longum affect colonic microbiology and fecal metabolites in health subjects? *The American Journal of Clinical Nutrition*, *59*(2), 428–432. https://doi.org/10.1093/ajcn/59.2.428

Bashiardes, S., Shapiro, H., Rozin, S., Shibolet, O., & Elinav, E. (2016). Non-alcoholic fatty liver and the gut microbiota. *Molecular Metabolism*, *5*(9), 782–794. https://doi.org/10.1016/j.molmet.2016.06.003

Bausch, D. G., Borchert, M., Grein, T., Roth, C., Swanepoel, R., Libande, M. L., … Rollin, P. E. (2003). Risk Factors for Marburg Hemorrhagic Fever, Democratic Republic of the Congo. *Emerging Infectious Diseases*, *9*(12), 1531–1537. https://doi.org/10.3201/eid0912.030355

Bekure, S., de Leeuw, P., Grandin, B., & Neate, P. (Series Ed.). (1991). *Maasai Herding: An analysis of the livestock production system of Maasai pastoralists in eastern*

Kajiado District, Kenya (p. 172). Addis Ababa, Ethiopia: International Livestock Centre for Africa.

Bello, M. G. D., Knight, R., Gilbert, J. A., & Blaser, M. J. (2018). Preserving microbial diversity. *Science*, *362*(6410), 33–34. https://doi.org/10.1126/science.aau8816

Beltrame, M. H., Rubel, M. A., & Tishkoff, S. A. (2016). Inferences of African evolutionary history from genomic data. *Current Opinion in Genetics & Development*, *41*, 159–166. https://doi.org/10.1016/j.gde.2016.10.002

Benno, Y., Endo, K., Mizutani, T., Namba, Y., Komori, T., & Mitsuoka, T. (1989). Comparison of fecal microflora of elderly persons in rural and urban areas of Japan. *Applied and Environmental Microbiology*, *55*(5), 1100–1105.

Berbesque, J. C., Marlowe, F. W., & Crittenden, A. N. (2011). Sex differences in Hadza eating frequency by food type. *American Journal of Human Biology: The Official Journal of the Human Biology Council*, *23*(3), 339–345. https://doi.org/10.1002/ajhb.21139

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., … Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics*, *74*(6), 1111–1120. https://doi.org/10.1086/421051

Bialonska, D., Ramnani, P., Kasimsetty, S. G., Muntha, K. R., Gibson, G. R., & Ferreira, D. (2010). The influence of pomegranate by-product and punicalagins on selected groups of human intestinal microbiota. *International Journal of Food Microbiology*, *140*(2–3), 175–182. https://doi.org/10.1016/j.ijfoodmicro.2010.03.038

Björkstén, B., Sepp, E., Julge, K., Voor, T., & Mikelsaar, M. (2001). Allergy development and the intestinal microflora during the first year of life. *The Journal of Allergy and Clinical Immunology, 108*(4), 516–520. https://doi.org/10.1067/mai.2001.118130

Blaser, M., Bork, P., Fraser, C., Knight, R., & Wang, J. (2013). The microbiome explored: Recent insights and future challenges. *Nature Reviews Microbiology, 11*(3), 213–217. https://doi.org/10.1038/nrmicro2973

Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q., Bukowski, R., Bell, J. T., … Clark, A. G. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0759-1

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolnick, D. I., Snowberg, L. K., Hirsch, P. E., Lauber, C. L., Org, E., Parks, B., … Svanbäck, R. (2014). Individual diet has sex-dependent effects on vertebrate gut microbiota. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms5500

Bonder, M. J., Kurilshikov, A., Tigchelaar, E. F., Mujagic, Z., Imhann, F., Vila, A. V., … Zhernakova, A. (2016). The effect of host genetics on the gut microbiome. *Nature Genetics*, *48*(11), 1407–1412. https://doi.org/10.1038/ng.3663

Bonder, M. J., Tigchelaar, E. F., Cai, X., Trynka, G., Cenit, M. C., Hrdlickova, B., … Zhernakova, A. (2016). The influence of a short-term gluten-free diet on the

human gut microbiome. *Genome Medicine*, *8*, 45. https://doi.org/10.1186/s13073-016-0295-y

Bougnoux, M.-E., Diogo, D., François, N., Sendid, B., Veirmeire, S., Colombel, J. F., … Poulain, D. (2006). Multilocus Sequence Typing Reveals Intrafamilial Transmission and Microevolutions of Candida albicans Isolates from the Human Digestive Tract. *Journal of Clinical Microbiology*, *44*(5), 1810–1820. https://doi.org/10.1128/JCM.44.5.1810-1820.2006

Bouhnik, Y., Flourie, B., Andrieux, C., Bisetti, N., Briet, F., & Rambaud, J. C. (1996). Effects of Bifidobacterium sp fermented milk ingested with or without inulin on colonic bifidobacteria and enzymatic activities in healthy humans. *European Journal of Clinical Nutrition*, *50*(4), 269–273.

Boyeldieu, P., Dimmendaal, G., Fleisch, A., Frajzyngier, Z., Güldemann, T., Nougayrol, P., … Vossen, R. (2008). *Problems of linguistic-historical reconstruction in Africa (SUGIA Sprache und Geschichte in Afrika)* (1st editions). Rüdiger Köppe.

Bräuer, G. (2008). The origin of modern anatomy: By speciation or intraspecific evolution? *Evolutionary Anthropology: Issues, News, and Reviews*, *17*(1), 22–37. https://doi.org/10.1002/evan.20157

Breitwieser, F. P., Baker, D. N., & Salzberg, S. L. (2018). KrakenUniq: Confident and fast metagenomics classification using unique k-mer counts. *Genome Biology*, *19*(1), 198. https://doi.org/10.1186/s13059-018-1568-0

Brenchley, J. M., Price, D. A., Schacker, T. W., Asher, T. E., Silvestri, G., Rao, S., … Douek, D. C. (2006). Microbial translocation is a cause of systemic immune

activation in chronic HIV infection. *Nature Medicine*, *12*(12), 1365–1371. https://doi.org/10.1038/nm1511

Brenner, D. M., Moeller, M. J., Chey, W. D., & Schoenfeld, P. S. (2009). The utility of probiotics in the treatment of irritable bowel syndrome: A systematic review. *The American Journal of Gastroenterology*, *104*(4), 1033–1049; quiz 1050. https://doi.org/10.1038/ajg.2009.25

Breton, G., Schlebusch, C. M., Lombard, M., Sjödin, P., Soodyall, H., & Jakobsson, M. (2014). Lactase persistence alleles reveal partial East African ancestry of southern African Khoe pastoralists. *Current Biology: CB*, *24*(8), 852–858. https://doi.org/10.1016/j.cub.2014.02.041

Broecker, F., Russo, G., Klumpp, J., & Moelling, K. (2017). Stable core virome despite variable microbiome after fecal transfer. *Gut Microbes*, *8*(3), 214–220. https://doi.org/10.1080/19490976.2016.1265196

Brumbaugh, D. E., Arruda, J., Robbins, K., Ir, D., Santorico, S. A., Robertson, C. E., & Frank, D. N. (2016). Mode of Delivery Determines Neonatal Pharyngeal Bacterial Composition and Early Intestinal Colonization. *Journal of Pediatric Gastroenterology and Nutrition*, *63*(3), 320–328. https://doi.org/10.1097/MPG.0000000000001124

Bryc, K., Auton, A., Nelson, M. R., Oksenberg, J. R., Hauser, S. L., Williams, S., … Bustamante, C. D. (2010). Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(2), 786–791. https://doi.org/10.1073/pnas.0909559107

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Bull-Otterson, L., Feng, W., Kirpich, I., Wang, Y., Qin, X., Liu, Y., … Barve, S. (2013). Metagenomic analyses of alcohol induced pathogenic alterations in the intestinal microbiome and the effect of Lactobacillus rhamnosus GG treatment. *PloS One*, *8*(1), e53028. https://doi.org/10.1371/journal.pone.0053028

Burgess, S. L., Buonomo, E., Carey, M., Cowardin, C., Naylor, C., Noor, Z., … Petri, W. A. (2014). Bone marrow dendritic cells from mice with an altered microbiota provide interleukin 17A-dependent protection against Entamoeba histolytica colitis. *MBio*, *5*(6), e01817. https://doi.org/10.1128/mBio.01817-14

Busby, G., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V., … Spencer, C. (2016). Admixture into and within sub-Saharan Africa. *BioRxiv*, 038406–038406. https://doi.org/10.1101/038406

Bzhalava, Z., Hultin, E., & Dillner, J. (2018). Extension of the viral ecology in humans using viral profile hidden Markov models. *PLOS ONE*, *13*(1), e0190938. https://doi.org/10.1371/journal.pone.0190938

Caesar, R., Tremaroli, V., Kovatcheva-Datchary, P., Cani, P. D., & Bäckhed, F. (2015). Crosstalk between Gut Microbiota and Dietary Lipids Aggravates WAT Inflammation through TLR Signaling. *Cell Metabolism*, *22*(4), 658–668. https://doi.org/10.1016/j.cmet.2015.07.026

Calattini, S., Betsem, E. B. A., Froment, A., Mauclère, P., Tortevoye, P., Schmitt, C., … Gessain, A. (2007). Simian foamy virus transmission from apes to humans, rural

Cameroon. *Emerging Infectious Diseases*, *13*(9), 1314–1320. https://doi.org/10.3201/eid1309.061162

Campbell, M. C., Ranciaro, A., Froment, A., Hirbo, J., Omar, S., Bodo, J.-M., … Tishkoff, S. A. (2012). Evolution of functionally diverse alleles associated with PTC bitter taste sensitivity in Africa. *Molecular Biology and Evolution*, *29*(4), 1141–1153. https://doi.org/10.1093/molbev/msr293

Campbell, M. C., & Tishkoff, S. A. (2008). African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics*, *9*, 403–433. https://doi.org/10.1146/annurev.genom.9.081307.164258

Campbell, M. C., & Tishkoff, S. A. (2010). The evolution of human genetic and phenotypic variation in Africa. *Current Biology: CB*, *20*(4), R166-173. https://doi.org/10.1016/j.cub.2009.11.050

Cani, P. D., Amar, J., Iglesias, M. A., Poggi, M., Knauf, C., Bastelica, D., … Burcelin, R. (2007). Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes*, *56*(7), 1761–1772. https://doi.org/10.2337/db06-1491

Cani, P. D., Bibiloni, R., Knauf, C., Waget, A., Neyrinck, A. M., Delzenne, N. M., & Burcelin, R. (2008). Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. *Diabetes*, *57*(6), 1470–1481. https://doi.org/10.2337/db07-1403

Cann, H. M., de Toma, C., Cazes, L., Legrand, M.-F., Morel, V., Piouffre, L., … Cavalli-Sforza, L. L. (2002). A human genome diversity cell line panel. *Science (New York, N.Y.)*, *296*(5566), 261–262.

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., & Knight, R. (2010). PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics (Oxford, England)*, *26*(2), 266–267. https://doi.org/10.1093/bioinformatics/btp636

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335–336. https://doi.org/10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., … 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*(Suppl 1), 4516–4522. https://doi.org/10.1073/pnas.1000080107

Carding, S. R., Davis, N., & Hoyles, L. (2017). Review article: The human intestinal virome in health and disease. *Alimentary Pharmacology & Therapeutics*, *46*(9), 800–815. https://doi.org/10.1111/apt.14280

Carvalho-Wells, A. L., Helmolz, K., Nodet, C., Molzer, C., Leonard, C., McKevith, B., … Tuohy, K. M. (2010). Determination of the in vivo prebiotic potential of a maize-based whole grain breakfast cereal: A human feeding study. *The British Journal of Nutrition*, *104*(9), 1353–1356. https://doi.org/10.1017/S0007114510002084

Castro-Mejía, J. L., Muhammed, M. K., Kot, W., Neve, H., Franz, C. M. A. P., Hansen, L. H., … Nielsen, D. S. (2015). Optimizing protocols for extraction of

bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome, 3*, 64. https://doi.org/10.1186/s40168-015-0131-4

Cavalli-Sforza, L. (1986). *African Pygmies*. New York: Academic Press.

CDC Press Releases. (2016, January 1). Retrieved October 30, 2018, from CDC website: https://www.cdc.gov/media/releases/2015/p0225-clostridium-difficile.html

Cenit, M. C., Olivares, M., Codoñer-Franch, P., & Sanz, Y. (2015). Intestinal Microbiota and Celiac Disease: Cause, Consequence or Co-Evolution? *Nutrients*, *7*(8), 6900–6923. https://doi.org/10.3390/nu7085314

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7. https://doi.org/10.1186/s13742-015-0047-8

Chappell, C. L., Darkoh, C., Shimmin, L., Farhana, N., Kim, D.-K., Okhuysen, P. C., & Hixson, J. (2016). Fecal Indole as a Biomarker of Susceptibility to Cryptosporidium Infection. *Infection and Immunity*, *84*(8), 2299–2306. https://doi.org/10.1128/IAI.00336-16

Chen, J., Bittinger, K., Charlson, E. S., Hoffmann, C., Lewis, J., Wu, G. D., … Li, H. (2012). Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics (Oxford, England)*, *28*(16), 2106–2113. https://doi.org/10.1093/bioinformatics/bts342

Chihanga, S., Haque, U., Chanda, E., Mosweunyane, T., Moakofhi, K., Jibril, H. B., … Glass, G. E. (2016). Malaria elimination in Botswana, 2012–2014: Achievements and challenges. *Parasites & Vectors*, *9*, 99. https://doi.org/10.1186/s13071-016-1382-z

Chinwe, E., & Agi, P. (2014). Urogenital schistosomiasis among Fulani pastoralists in Rivers state, Nigeria. *13th International Congress of Parasitology*. Presented at the World Federation of Parasitologists.

Chitsulo, L., Engels, D., Montresor, A., & Savioli, L. (2000). The global status of schistosomiasis and its control. *Acta Tropica, 77*(1), 41–51.

Chong, C. W., Ahmad, A. F., Lim, Y. A. L., Teh, C. S. J., Yap, I. K. S., Lee, S. C., … Chua, K. H. (2015). Effect of ethnicity and socioeconomic variation to the gut microbiota composition among pre-adolescent in Malaysia. *Scientific Reports, 5*. https://doi.org/10.1038/srep13338

Christensen, D. l., Faurholt-Jepsen, D., Boit, M. k., Mwaniki, D. l., Kilonzo, B., Tetens, I., … Brage, S. (2012). Cardiorespiratory fitness and physical activity in Luo, Kamba, and Maasai of rural Kenya. *American Journal of Human Biology, 24*(6), 723–729. https://doi.org/10.1002/ajhb.22303

Clark, J. D., Beyene, Y., WoldeGabriel, G., Hart, W. K., Renne, P. R., Gilbert, H., … White, T. D. (2003). Stratigraphic, chronological and behavioural contexts of Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature*, *423*(6941), 747–752. https://doi.org/10.1038/nature01670

Clarke, E. L., Taylor, L. J., Zhao, C., Connell, A., Lee, J.-J., Fett, B., … Bittinger, K. (2019). Sunbeam: An extensible pipeline for analyzing metagenomic sequencing experiments. *Microbiome*, *7*(1), 46. https://doi.org/10.1186/s40168-019-0658-x

Clayton, J. B., Gomez, A., Amato, K., Knights, D., Travis, D. A., Blekhman, R., … Johnson, T. J. (2018). The gut microbiome of nonhuman primates: Lessons in

ecology and evolution. *American Journal of Primatology*, *80*(6), e22867. https://doi.org/10.1002/ajp.22867

Clemente, J. C., Pehrsson, E. C., Blaser, M. J., Sandhu, K., Gao, Z., Wang, B., … Dominguez-Bello, M. G. (2015). The microbiome of uncontacted Amerindians. *Science Advances*, *1*(3), e1500183–e1500183. https://doi.org/10.1126/sciadv.1500183

Clemente, Jose C., Ursell, L. K., Parfrey, L. W., & Knight, R. (2012). The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell*, *148*(6), 1258–1270. https://doi.org/10.1016/j.cell.2012.01.035

Clooney, A. G., Fouhy, F., Sleator, R. D., Driscoll, A. O., Stanton, C., Cotter, P. D., & Claesson, M. J. (2016). Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis. *PLOS ONE*, *11*(2), e0148028. https://doi.org/10.1371/journal.pone.0148028

Coenye, T., & Vandamme, P. (2003). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiology Letters*, *228*(1), 45–49.

Coghlan, A., Tyagi, R., Cotton, J. A., Holroyd, N., Rosa, B. A., Tsai, I. J., … International Helminth Genomes Consortium. (2019). Comparative genomics of the major parasitic worms. *Nature Genetics*, *51*(1), 163–174. https://doi.org/10.1038/s41588-018-0262-1

Cohen, D. (2002). *Human capital and the HIV epidemic in sub-Saharan Africa*. Geneva: ILO Programme on HIV/AIDS and the World of Work.

Collado, M. C., Donat, E., Ribes-Koninckx, C., Calabuig, M., & Sanz, Y. (2009). Specific duodenal and faecal bacterial groups associated with paediatric coeliac disease. *Journal of Clinical Pathology, 62*(3), 264–269. https://doi.org/10.1136/jcp.2008.061366

Collado, M. Carmen, Surono, I. S., Meriluoto, J., & Salminen, S. (2007). Potential probiotic characteristics of Lactobacillus and Enterococcus strains isolated from traditional dadih fermented milk against pathogen intestinal colonization. *Journal of Food Protection*, *70*(3), 700–705.

Consortium, T. H. M. P. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. https://doi.org/10.1038/nature11234

Cooper, P., Walker, A. W., Reyes, J., Chico, M., Salter, S. J., Vaca, M., & Parkhill, J. (2013). Patent Human Infections with the Whipworm, Trichuris trichiura, Are Not Associated with Alterations in the Faecal Microbiota. *PLOS ONE*, *8*(10), e76573. https://doi.org/10.1371/journal.pone.0076573

Costabile, A., Fava, F., Röytiö, H., Forssten, S. D., Olli, K., Klievink, J., … Walton, G. E. (2012). Impact of polydextrose on the faecal microbiota: A double-blind, crossover, placebo-controlled feeding study in healthy human subjects. *The British Journal of Nutrition*, *108*(3), 471–481. https://doi.org/10.1017/S0007114511005782

Costabile, A., Klinder, A., Fava, F., Napolitano, A., Fogliano, V., Leonard, C., … Tuohy, K. M. (2008). Whole-grain wheat breakfast cereal has a prebiotic effect on the human gut microbiota: A double-blind, placebo-controlled, crossover study. *The*

British Journal of Nutrition, *99*(1), 110–120. https://doi.org/10.1017/S0007114507793923

Cotillard, A., Kennedy, S. P., Kong, L. C., Prifti, E., Pons, N., Le Chatelier, E., … Ehrlich, S. D. (2013). Dietary intervention impact on gut microbial gene richness. *Nature*, *500*(7464), 585–588. https://doi.org/10.1038/nature12480

Cox, L. M., Yamanishi, S., Sohn, J., Alekseyenko, A. V., Leung, J. M., Cho, I., … Blaser, M. J. (2014). Altering the Intestinal Microbiota during a Critical Developmental Window Has Lasting Metabolic Consequences. *Cell*, *158*(4), 705–721. https://doi.org/10.1016/j.cell.2014.05.052

Crawford, N. G., Kelly, D. E., Hansen, M. E. B., Beltrame, M. H., Fan, S., Bowman, S. L., … Tishkoff, S. (2017). Loci associated with skin pigmentation identified in African populations. *Science*, eaan8433. https://doi.org/10.1126/science.aan8433

Cuervo, A., Hevia, A., López, P., Suárez, A., Diaz, C., Sánchez, B., … González, S. (2016). Phenolic compounds from red wine and coffee are associated with specific intestinal microorganisms in allergic subjects. *Food & Function*, *7*(1), 104–109. https://doi.org/10.1039/c5fo00853k

Cuervo, A., Valdés, L., Salazar, N., de los Reyes-Gavilán, C. G., Ruas-Madiedo, P., Gueimonde, M., & González, S. (2014). Pilot study of diet and microbiota: Interactive associations of fibers and polyphenols with human intestinal bacteria. *Journal of Agricultural and Food Chemistry*, *62*(23), 5330–5336. https://doi.org/10.1021/jf501546a

Cueva, C., Sánchez-Patán, F., Monagas, M., Walton, G. E., Gibson, G. R., Martín-Álvarez, P. J., … Moreno-Arribas, M. V. (2013). In vitro fermentation of grape

seed flavan-3-ol fractions by human faecal microbiota: Changes in microbial groups and phenolic metabolites. *FEMS Microbiology Ecology*, *83*(3), 792–805. https://doi.org/10.1111/1574-6941.12037

Dalmasso, M., Hill, C., & Ross, R. P. (2014). Exploiting gut bacteriophages for human health. *Trends in Microbiology*, *22*(7), 399–405. https://doi.org/10.1016/j.tim.2014.02.010

Dao, M. C., Everard, A., Aron-Wisnewsky, J., Sokolovska, N., Prifti, E., Verger, E. O., … Clément, K. (2016). Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: Relationship with gut microbiome richness and ecology. *Gut*, *65*(3), 426–436. https://doi.org/10.1136/gutjnl-2014-308778

Davenport, E. R., Mizrahi-Man, O., Michelini, K., Barreiro, L. B., Ober, C., & Gilad, Y. (2014). Seasonal Variation in Human Gut Microbiome Composition. *PLOS ONE*, *9*(3), e90731. https://doi.org/10.1371/journal.pone.0090731

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., … Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559–563. https://doi.org/10.1038/nature12820

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., … 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*(33), 14691–14696. https://doi.org/10.1073/pnas.1005963107

de Groot, P. F., Frissen, M. N., de Clercq, N. C., & Nieuwdorp, M. (2017). Fecal microbiota transplantation in metabolic syndrome: History, present and future. *Gut Microbes*, *8*(3), 253–267. https://doi.org/10.1080/19490976.2017.1293224

De Palma, G., Nadal, I., Collado, M. C., & Sanz, Y. (2009). Effects of a gluten-free diet on gut microbiota and immune function in healthy adult human subjects. *The British Journal of Nutrition*, *102*(8), 1154–1160. https://doi.org/10.1017/S0007114509371767

De Palma, G., Nadal, I., Medina, M., Donat, E., Ribes-Koninckx, C., Calabuig, M., & Sanz, Y. (2010). Intestinal dysbiosis and reduced immunoglobulin-coated bacteria associated with coeliac disease in children. *BMC Microbiology*, *10*, 63. https://doi.org/10.1186/1471-2180-10-63

Del Chierico, F., Vernocchi, P., Dallapiccola, B., & Putignani, L. (2014). Mediterranean diet and health: Food effects on gut microbiota and disease control. *International Journal of Molecular Sciences*, *15*(7), 11678–11699. https://doi.org/10.3390/ijms150711678

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., … Andersen, G. L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.*, *72*(7), 5069–5072. https://doi.org/10.1128/AEM.03006-05

Deschasaux, M., Bouter, K. E., Prodan, A., Levin, E., Groen, A. K., Herrema, H., … Nieuwdorp, M. (2018). Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. *Nature Medicine*, *24*(10), 1526. https://doi.org/10.1038/s41591-018-0160-1

Devkota, S., Wang, Y., Musch, M. W., Leone, V., Fehlner-Peach, H., Nadimpalli, A., … Chang, E. B. (2012). Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. *Nature*, *487*(7405), 104–108. https://doi.org/10.1038/nature11225

Dheilly, N. M., Bolnick, D., Bordenstein, S., Brindley, P. J., Figuères, C., Holmes, E. C., … Rosario, K. (2017). Parasite Microbiome Project: Systematic Investigation of Microbiome Dynamics within and across Parasite-Host Interactions. *MSystems*, *2*(4). https://doi.org/10.1128/mSystems.00050-17

Dicksved, J., Flöistrup, H., Bergström, A., Rosenquist, M., Pershagen, G., Scheynius, A., … Jansson, J. K. (2007). Molecular Fingerprinting of the Fecal Microbiota of Children Raised According to Different Lifestyles. *Applied and Environmental Microbiology*, *73*(7), 2284–2289. https://doi.org/10.1128/AEM.02223-06

Dillon, S., Lee, E., Kotter, C., Austin, G., Dong, Z., Hecht, D., … Wilson, C. (2014). An altered intestinal mucosal microbiome in HIV-1 infection is associated with mucosal and systemic immune activation and endotoxemia. *Mucosal Immunology*, *7*(4), 983–994. https://doi.org/10.1038/mi.2013.116

Dinh, D. M., Volpe, G. E., Duffalo, C., Bhalchandra, S., Tai, A. K., Kane, A. V., … Ward, H. D. (2015). Intestinal Microbiota, Microbial Translocation, and Systemic Inflammation in Chronic HIV Infection. *The Journal of Infectious Diseases*, *211*(1), 19–27. https://doi.org/10.1093/infdis/jiu409

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of*

the National Academy of Sciences of the United States of America, *107*(26), 11971–11975. https://doi.org/10.1073/pnas.1002601107

Dominianni, C., Sinha, R., Goedert, J. J., Pei, Z., Yang, L., Hayes, R. B., & Ahn, J. (2015). Sex, body mass index, and dietary fiber intake influence the human gut microbiome. *PloS One*, *10*(4), e0124599. https://doi.org/10.1371/journal.pone.0124599

Drasar, B. S., Crowther, J. S., Goddard, P., Hawksworth, G., Hill, M. J., Peach, S., … Renwick, A. (1973). The relation between diet and the gut microflora in man. *The Proceedings of the Nutrition Society*, *32*(2), 49–52.

Dridi, B., Fardeau, M.-L., Ollivier, B., Raoult, D., & Drancourt, M. (2011). The antimicrobial resistance pattern of cultured human methanogens reflects the unique phylogenetic position of archaea. *The Journal of Antimicrobial Chemotherapy*, *66*(9), 2038–2044. https://doi.org/10.1093/jac/dkr251

Druart, C., Dewulf, E. M., Cani, P. D., Neyrinck, A. M., Thissen, J.-P., & Delzenne, N. M. (2014). Gut microbial metabolites of polyunsaturated fatty acids correlate with specific fecal bacteria and serum markers of metabolic syndrome in obese women. *Lipids*, *49*(4), 397–402. https://doi.org/10.1007/s11745-014-3881-z

Dutilh, B. E., Cassman, N., McNair, K., Sanchez, S. E., Silva, G. G. Z., Boling, L., … Edwards, R. A. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature Communications*, *5*, 4498. https://doi.org/10.1038/ncomms5498

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., … Relman, D. A. (2005). Diversity of the Human Intestinal Microbial Flora. *Science*

(New York, N.Y.), *308*(5728), 1635–1638. https://doi.org/10.1126/science.1110591

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Eeckhaut, V., Machiels, K., Perrier, C., Romero, C., Maes, S., Flahou, B., … Immerseel, F. V. (2013). Butyricicoccus pullicaecorum in inflammatory bowel disease. *Gut*, *62*(12), 1745–1752. https://doi.org/10.1136/gutjnl-2012-303611

Eggimann, P., Garbino, J., & Pittet, D. (2003). Epidemiology of Candida species infections in critically ill non-immunosuppressed patients. *The Lancet. Infectious Diseases*, *3*(11), 685–702.

Ehret, C. (2000). *Language and History, in "African languages: An introduction."* Cambridge [England]; New York: Cambridge University Press.

Eid, N., Enani, S., Walton, G., Corona, G., Costabile, A., Gibson, G., … Spencer, J. P. E. (2014). The impact of date palm fruits and their component polyphenols, on gut microbial ecology, bacterial metabolites and colon cancer cell proliferation. *Journal of Nutritional Science*, *3*. https://doi.org/10.1017/jns.2014.16

Elderman, M., Sovran, B., Hugenholtz, F., Graversen, K., Huijskes, M., Houtsma, E., … Faas, M. (2017). The effect of age on the intestinal mucus thickness, microbiota composition and immunity in relation to sex in mice. *PLOS ONE*, *12*(9), e0184274. https://doi.org/10.1371/journal.pone.0184274

Elliott, D. E., & Weinstock, J. V. (2012). Helminth-host immunological interactions: Prevention and control of immune-mediated diseases. *Annals of the New York*

Academy of Sciences, *1247*, 83–96. https://doi.org/10.1111/j.1749-6632.2011.06292.x

Enattah, N. S., Trudeau, A., Pimenoff, V., Maiuri, L., Auricchio, S., Greco, L., … Peltonen, L. (2007). Evidence of Still-Ongoing Convergence Evolution of the Lactase Persistence T-13910 Alleles in Humans. *American Journal of Human Genetics*, *81*(3), 615–625.

Etienne, R. S. (2005). A new sampling formula for neutral biodiversity. *Ecology Letters*, *8*(3), 253–260. https://doi.org/10.1111/j.1461-0248.2004.00717.x

Fagny, M., Patin, E., MacIsaac, J. L., Rotival, M., Flutre, T., Jones, M. J., … Quintana-Murci, L. (2015). The epigenomic landscape of African rainforest hunter-gatherers and farmers. *Nature Communications*, *6*, 10047. https://doi.org/10.1038/ncomms10047

Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., … Gordon, J. I. (2013). The long-term stability of the human gut microbiota. *Science (New York, N.Y.)*, *341*(6141), 1237439. https://doi.org/10.1126/science.1237439

Fall, T., Lundholm, C., Örtqvist, A. K., Fall, K., Fang, F., Hedhammar, Å., … Almqvist, C. (2015). Early Exposure to Dogs and Farm Animals and the Risk of Childhood Asthma. *JAMA Pediatrics*, *169*(11), e153219. https://doi.org/10.1001/jamapediatrics.2015.3219

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., … Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science*, *352*(6285), 560–564. https://doi.org/10.1126/science.aad3503

Fan, S., Hansen, M. E. B., Lo, Y., & Tishkoff, S. A. (2016). Going global by adapting local: A review of reent human adaptation. *Science, 354*(6308). https://doi.org/10.1126/science.aaf5098

Fan, S., Kelly, D. E., Beltrame, M. H., Hansen, M. E. B., Mallick, S., Ranciaro, A., … Tishkoff, S. A. (2019). African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biology, 20*(1), 82. https://doi.org/10.1186/s13059-019-1679-2

Fava, F., Gitau, R., Griffin, B. A., Gibson, G. R., Tuohy, K. M., & Lovegrove, J. A. (2013). The type and quantity of dietary fat and carbohydrate alter faecal microbiome and short-chain fatty acid excretion in a metabolic syndrome 'at-risk' population. *International Journal of Obesity*, *37*(2), 216–223. https://doi.org/10.1038/ijo.2012.33

Fernandes, J., Su, W., Rahat-Rozenbloom, S., Wolever, T. M. S., & Comelli, E. M. (2014). Adiposity, gut microbiota and faecal short chain fatty acids are linked in adult humans. *Nutrition & Diabetes*, *4*(6), e121. https://doi.org/10.1038/nutd.2014.23

Filippis, F. D., Pellegrini, N., Vannini, L., Jeffery, I. B., Storia, A. L., Laghi, L., … Ercolini, D. (2016). High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut, 65*(11), 1812–1821. https://doi.org/10.1136/gutjnl-2015-309957

Flickinger, E. A., Hatch, T. F., Wofford, R. C., Grieshop, C. M., Murray, S. M., & Fahey, G. C. (2002). In vitro fermentation properties of selected fructooligosaccharide-containing vegetables and in vivo colonic microbial populations are affected by

the diets of healthy human infants. *The Journal of Nutrition*, *132*(8), 2188–2194. https://doi.org/10.1093/jn/132.8.2188

Foligné, B., Parayre, S., Cheddani, R., Famelart, M.-H., Madec, M.-N., Plé, C., … Deutsch, S.-M. (2016). Immunomodulation properties of multi-species fermented milks. *Food Microbiology*, *53*(Pt A), 60–69. https://doi.org/10.1016/j.fm.2015.04.002

Fotedar, R., Stark, D., Beebe, N., Marriott, D., Ellis, J., & Harkness, J. (2007). Laboratory Diagnostic Techniques for Entamoeba Species. *Clinical Microbiology Reviews*, *20*(3), 511–532. https://doi.org/10.1128/CMR.00004-07

Francavilla, R., Calasso, M., Calace, L., Siragusa, S., Ndagijimana, M., Vernocchi, P., … De Angelis, M. (2012). Effect of lactose on gut microbiota and metabolome of infants with cow's milk allergy. *Pediatric Allergy and Immunology: Official Publication of the European Society of Pediatric Allergy and Immunology*, *23*(5), 420–427. https://doi.org/10.1111/j.1399-3038.2012.01286.x

François, I. E. J. A., Lescroart, O., Veraverbeke, W. S., Marzorati, M., Possemiers, S., Hamer, H., … Broekaert, W. F. (2014). Effects of wheat bran extract containing arabinoxylan oligosaccharides on gastrointestinal parameters in healthy preadolescent children. *Journal of Pediatric Gastroenterology and Nutrition*, *58*(5), 647–653. https://doi.org/10.1097/MPG.0000000000000285

Frank, D. N., St. Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N., & Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National*

Academy of Sciences of the United States of America, *104*(34), 13780–13785. https://doi.org/10.1073/pnas.0706625104

Froment, A. (2014). Human Biology and Health of African Rainforest Inhabitants. In *Hunter Gatherers of the Congo Basin* (pp. 117–164). New Jersey: Transaction Publishers.

Gagneux, S. (2012). Host–pathogen coevolution in human tuberculosis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1590), 850–859. https://doi.org/10.1098/rstb.2011.0316

Galperin, M. Y., Brover, V., Tolstoy, I., & Yutin, N. (2016). Phylogenomic analysis of the family Peptostreptococcaceae (Clostridium cluster XI) and proposal for reclassification of Clostridium litorale (Fendrich et al. 1991) and Eubacterium acidaminophilum (Zindel et al. 1989) as Peptoclostridium litorale gen. Nov. Comb. Nov. And Peptoclostridium acidaminophilum comb. Nov. *International Journal of Systematic and Evolutionary Microbiology*, *66*(12), 5506–5513. https://doi.org/10.1099/ijsem.0.001548

García-Albiach, R., Pozuelo de Felipe, M. J., José, M., de Felipe, P., Angulo, S., Morosini, M.-I., … del Campo, R. (2008). Molecular analysis of yogurt containing Lactobacillus delbrueckii subsp. Bulgaricus and Streptococcus thermophilus in human intestinal microbiota. *The American Journal of Clinical Nutrition*, *87*(1), 91–96. https://doi.org/10.1093/ajcn/87.1.91

Gazanion, E., & Vergnes, B. (2018). Protozoan Parasite Auxotrophies and Metabolic Dependencies. *Metabolic Interaction in Infection*, 351–375. https://doi.org/10.1007/978-3-319-74932-7_9

Genovese, G., Friedman, D. J., Ross, M. D., Lecordier, L., Uzureau, P., Freedman, B. I., … Pollak, M. R. (2010). Association of Trypanolytic ApoL1 Variants with Kidney Disease in African Americans. *Science, 329*(5993), 841–845. https://doi.org/10.1126/science.1193032

Ghannoum, M. A., Jurevic, R. J., Mukherjee, P. K., Cui, F., Sikaroodi, M., Naqvi, A., & Gillevet, P. M. (2010). Characterization of the Oral Fungal Microbiome (Mycobiome) in Healthy Individuals. *PLoS Pathog, 6*(1), e1000713. https://doi.org/10.1371/journal.ppat.1000713

Giacomin, P., Zakrzewski, M., Croese, J., Su, X., Sotillo, J., McCann, L., … Cantacessi, C. (2015). Experimental hookworm infection and escalating gluten challenges are associated with increased microbial richness in celiac subjects. *Scientific Reports, 5*. https://doi.org/10.1038/srep13797

Giacomin, P., Zakrzewski, M., Jenkins, T. P., Su, X., Al-Hallaf, R., Croese, J., … Cantacessi, C. (2016). Changes in duodenal tissue-associated microbiota following hookworm infection and consecutive gluten challenges in humans with coeliac disease. *Scientific Reports, 6,* 36797. https://doi.org/10.1038/srep36797

Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H., … Nussbaum, R. L. (2003). The International HapMap Project. *Nature, 426*(6968), 789–796. https://doi.org/10.1038/nature02168

Gilbert, S. F. (2014). A holobiont birth narrative: The epigenetic transmission of the human microbiome. *Frontiers in Genetics, 5*. https://doi.org/10.3389/fgene.2014.00282

Gilchrist, C. A., Petri, S. E., Schneider, B. N., Reichman, D. J., Jiang, N., Begum, S., … Petri, W. A. (2016). Role of the Gut Microbiota of Children in Diarrhea Due to the Protozoan Parasite Entamoeba histolytica. *The Journal of Infectious Diseases*, *213*(10), 1579–1585. https://doi.org/10.1093/infdis/jiv772

Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., … Nelson, K. E. (2006). Metagenomic Analysis of the Human Distal Gut Microbiome. *Science (New York, N.Y.)*, *312*(5778), 1355–1359. https://doi.org/10.1126/science.1124234

Global Tuberculosis Control: WHO Report 2010. (2010). World Health Organization.

Goedert, J. J. (2016). Effects of HIV, Immune Deficiency, and Confounding on the Distal Gut Microbiota. *EBioMedicine*, *5*, 14–15. https://doi.org/10.1016/j.ebiom.2016.01.034

Goldin, B. R., & Gorbach, S. L. (1980). Effect of Lactobacillus acidophilus dietary supplements on 1,2-dimethylhydrazine dihydrochloride-induced intestinal cancer in rats. *Journal of the National Cancer Institute*, *64*(2), 263–265.

Goldin, B. R., Gualtieri, L. J., & Moore, R. P. (1996). The effect of Lactobacillus GG on the initiation and promotion of DMH-induced intestinal tumors in the rat. *Nutrition and Cancer*, *25*(2), 197–204. https://doi.org/10.1080/01635589609514442

Gomez, A., Petrzelkova, K. J., Burns, M. B., Yeoman, C. J., Amato, K. R., Vlckova, K., … Blekhman, R. (2016). Gut Microbiome of Coexisting BaAka Pygmies and Bantu Reflects Gradients of Traditional Subsistence Patterns. *Cell Reports*, *0*(0). https://doi.org/10.1016/j.celrep.2016.02.013

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., … Ley, R. E. (2016). Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host & Microbe*, *19*(5), 731–743. https://doi.org/10.1016/j.chom.2016.04.017

Goodrich, J. K., Davenport, E. R., Clark, A. G., & Ley, R. E. (2017). The Relationship Between the Human Genome and Microbiome Comes into View. *Annual Review of Genetics*, *51*, 413–433. https://doi.org/10.1146/annurev-genet-110711-155532

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., … Ley, R. E. (2014). Human Genetics Shape the Gut Microbiome. *Cell*, *159*(4), 789–799. https://doi.org/10.1016/j.cell.2014.09.053

Goossens, D. a. M., Jonkers, D. M. a. E., Russel, M. G. V. M., Stobberingh, E. E., & Stockbrügger, R. W. (2006). The effect of a probiotic drink with Lactobacillus plantarum 299v on the bacterial composition in faeces and mucosal biopsies of rectum and ascending colon. *Alimentary Pharmacology & Therapeutics*, *23*(2), 255–263. https://doi.org/10.1111/j.1365-2036.2006.02749.x

Gopalan, S., Carja, O., Fagny, M., Patin, E., Myrick, J. W., McEwen, L., … Henn, B. M. (2016). Trends in DNA methylation with age replicate across diverse human populations. *BioRxiv*, 073171. https://doi.org/10.1101/073171

Gopalan, S., Carja, O., Fagny, M., Patin, E., Myrick, J. W., McEwen, L. M., … Henn, B. M. (2017). Trends in DNA Methylation with Age Replicate Across Diverse Human Populations. *Genetics*, *206*(3), 1659–1674. https://doi.org/10.1534/genetics.116.195594

Gori, A., Rizzardini, G., Van't Land, B., Amor, K. B., van Schaik, J., Torti, C., … Clerici, M. (2011). Specific prebiotics modulate gut microbiota and immune activation in HAART-naive HIV-infected adults: Results of the "COPA" pilot randomized trial. *Mucosal Immunology*, *4*(5), 554–563. https://doi.org/10.1038/mi.2011.15

Gosalbes, M. J., Durbán, A., Pignatelli, M., Abellan, J. J., Jiménez-Hernández, N., Pérez-Cobas, A. E., … Moya, A. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PloS One*, *6*(3), e17447. https://doi.org/10.1371/journal.pone.0017447

Grammatikos, A. P. (2008). The genetic and environmental basis of atopic diseases. *Annals of Medicine*, *40*(7), 482–495. https://doi.org/10.1080/07853890802082096

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Griffith, D. M., Veech, J. A., & Marsh, C. J. (2016). cooccur: Probabilistic Species Co-Occurrence Analysis in R. *Journal of Statistical Software, Code Snippets*, *69*(2), 1–17. https://doi.org/10.18637/jss.v069.c02

Grivetti, L. E. (1978). Nutritional success in a semi-arid land: Examination of Tswana agro-pastoralists of the eastern Kalahari, Botswana. *The American Journal of Clinical Nutrition*, *31*(7), 1204–1220.

Guarner, F., & Malagelada, J.-R. (2003). Role of bacteria in experimental colitis. *Best Practice & Research. Clinical Gastroenterology*, *17*(5), 793–804.

Gupta, V. K., Paul, S., & Dutta, C. (2017). Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome Composition and Diversity. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.01162

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., … Sandhu, M. S. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature*, *517*(7534), 327–332. https://doi.org/10.1038/nature13997

Halmos, E. P., Christophersen, C. T., Bird, A. R., Shepherd, S. J., Gibson, P. R., & Muir, J. G. (2015). Diets that differ in their FODMAP content alter the colonic luminal microenvironment. *Gut*, *64*(1), 93–100. https://doi.org/10.1136/gutjnl-2014-307264

Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J., & Knight, R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods*, *5*(3), 235–237. https://doi.org/10.1038/nmeth.1184

Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C., & Wall, J. D. (2011). Genetic evidence for archaic admixture in Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(37), 15123–15128. https://doi.org/10.1073/pnas.1109300108

Hanevik, K., Wensaas, K.-A., Rortveit, G., Eide, G. E., Mørch, K., & Langeland, N. (2014). Irritable bowel syndrome and chronic fatigue 6 years after giardia infection: A controlled prospective cohort study. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *59*(10), 1394–1400. https://doi.org/10.1093/cid/ciu629

Hansen, M. E. B., Rubel, M. A., Bailey, A. G., Ranciaro, A., Thompson, S. R., Campbell, M. C., … Tishkoff, S. A. (2019). Population structure of human gut bacteria in a diverse cohort from rural Tanzania and Botswana. *Genome Biology*, *20*(1), 16. https://doi.org/10.1186/s13059-018-1616-9

Haro, C., Rangel-Zúñiga, O. A., Alcalá-Díaz, J. F., Gómez-Delgado, F., Pérez-Martínez, P., Delgado-Lista, J., … Camargo, A. (2016). Intestinal Microbiota Is Influenced by Gender and Body Mass Index. *PLOS ONE*, *11*(5), e0154090. https://doi.org/10.1371/journal.pone.0154090

Harrison, C. A., & Taren, D. (2017). How poverty affects diet to shape the microbiota and chronic disease. *Nature Reviews Immunology*. https://doi.org/10.1038/nri.2017.121

Harvati, K., Stringer, C., Grün, R., Aubert, M., Allsworth-Jones, P., & Folorunso, C. A. (2011). The Later Stone Age Calvaria from Iwo Eleru, Nigeria: Morphology and Chronology. *PLOS ONE*, *6*(9), e24024. https://doi.org/10.1371/journal.pone.0024024

Hassan, H. Y., van Erp, A., Jaeger, M., Tahir, H., Oosting, M., Joosten, L. A. B., & Netea, M. G. (2016). Genetic diversity of lactase persistence in East African populations. *BMC Research Notes*, *9*. https://doi.org/10.1186/s13104-015-1833-1

Hayashi, H., Sakamoto, M., Kitahara, M., & Benno, Y. (2003). Molecular analysis of fecal microbiota in elderly individuals using 16S rDNA library and T-RFLP. *Microbiology and Immunology*, *47*(8), 557–570.

He, T., Priebe, M. G., Zhong, Y., Huang, C., Harmsen, H. J. M., Raangs, G. C., … Vonk, R. J. (2008). Effects of yogurt and bifidobacteria supplementation on the colonic

microbiota in lactose-intolerant subjects. *Journal of Applied Microbiology*, *104*(2), 595–604. https://doi.org/10.1111/j.1365-2672.2007.03579.x

He, Y., Wu, W., Zheng, H.-M., Li, P., McDonald, D., Sheng, H.-F., … Zhou, H.-W. (2018). Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nature Medicine*, *24*(10), 1532. https://doi.org/10.1038/s41591-018-0164-x

Hehemann, J.-H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., & Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature*, *464*(7290), 908–912. https://doi.org/10.1038/nature08937

Heine, B., & Nurse, D. (2000). *African languages: An introduction*. Cambridge University Press.

Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D., & Myers, S. (2014). A Genetic Atlas of Human Admixture History. *Science*, *343*(6172), 747–751. https://doi.org/10.1126/science.1243518

Henao-Restrepo, A. M., Camacho, A., Longini, I. M., Watson, C. H., Edmunds, W. J., Egger, M., … Kieny, M.-P. (2017). Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: Final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!). *Lancet (London, England)*, *389*(10068), 505. https://doi.org/10.1016/S0140-6736(16)32621-6

Henckaerts, L., Figueroa, C., Vermeire, S., & Sans, M. (2008). The role of genetics in inflammatory bowel disease. *Current Drug Targets*, *9*(5), 361–368.

Hirbo, J., Ranciaro, A., & Tishkoff, S. (2012). *Population structure and migration in Africa: Correlations between archeological, linguistic, and genetic data. In Causes and Consequences of Migration. Edited by Campbell B and Crawford MH*. Cambridge University Press.

Hitchcock, R. (2002). "We Are the First People": Land, Natural Resources, and Identity in the Central Kalahari, Botswana. *Journal of Southern African Studies*, *28*(4), 797–824.

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G. D., … Bushman, F. D. (2013). Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *PLoS ONE*, *8*(6), e66019. https://doi.org/10.1371/journal.pone.0066019

Hooper, L. V., Littman, D. R., & Macpherson, A. J. (2012). Interactions between the microbiota and the immune system. *Science (New York, N.Y.)*, *336*(6086), 1268–1273. https://doi.org/10.1126/science.1223490

Hotez, P., Brindley, P., Bethony, J., King, C., Pearce, E., & Jacobson, J. (2008). Helminth infections: The great neglected tropical diseases. *The Journal of Clinical Investigation*, *118*, 1311–1321.

Hotez, P. J. (2014). Aboriginal Populations and Their Neglected Tropical Diseases. *PLoS Negl Trop Dis*, *8*(1), e2286. https://doi.org/10.1371/journal.pntd.0002286

Hotez, P. J., Molyneux, D. H., Fenwick, A., Ottesen, E., Ehrlich Sachs, S., & Sachs, J. D. (2006). Incorporating a Rapid-Impact Package for Neglected Tropical Diseases with Programs for HIV/AIDS, Tuberculosis, and Malaria. *PLoS Med*, *3*(5), e102. https://doi.org/10.1371/journal.pmed.0030102

Hoyles, L., McCartney, A. L., Neve, H., Gibson, G. R., Sanderson, J. D., Heller, K. J., & van Sinderen, D. (2014). Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Research in Microbiology*, *165*(10), 803–812. https://doi.org/10.1016/j.resmic.2014.10.006

Hsieh, P., Veeramah, K. R., Lachance, J., Tishkoff, S. A., Wall, J. D., Hammer, M. F., & Gutenkunst, R. N. (2016). Whole-genome sequence analyses of Western Central African Pygmy hunter-gatherers reveal a complex demographic history and identify candidate genes under positive natural selection. *Genome Research*. https://doi.org/10.1101/gr.192971.115

Hsieh, P., Woerner, A. E., Wall, J. D., Lachance, J., Tishkoff, S. A., Gutenkunst, R. N., & Hammer, M. F. (2016). Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies. *Genome Research*. https://doi.org/10.1101/gr.196634.115

Huang, L., & Appleton, J. A. (2016). Eosinophils in helminth infection: Defenders and dupes. *Trends in Parasitology*, *32*(10), 798–807. https://doi.org/10.1016/j.pt.2016.05.004

Hublin, J.-J., Ben-Ncer, A., Bailey, S. E., Freidline, S. E., Neubauer, S., Skinner, M. M., … Gunz, P. (2017). New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature*, *546*(7657), 289–292. https://doi.org/10.1038/nature22336

Human Microbiome Project Consortium. (2012). A framework for human microbiome research. *Nature*, *486*(7402), 215–221. https://doi.org/10.1038/nature11209

Humen, M. A., De Antoni, G. L., Benyacoub, J., Costas, M. E., Cardozo, M. I., Kozubsky, L., … Pérez, P. F. (2005). Lactobacillus johnsonii La1 Antagonizes Giardia intestinalis In Vivo. *Infection and Immunity*, *73*(2), 1265–1269. https://doi.org/10.1128/IAI.73.2.1265-1269.2005

Iebba, V., Santangelo, F., Totino, V., Pantanella, F., Monsia, A., Di Cristanziano, V., … D'Alfonso, R. (2016). Gut microbiota related to Giardia duodenalis, Entamoeba spp. And Blastocystis hominis infections in humans from Côte d'Ivoire. *Journal of Infection in Developing Countries*, *10*(9), 1035–1041. https://doi.org/10.3855/jidc.8179

Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314. https://doi.org/10.1080/10618600.1996.10474713

Ikeya, K. (2001). Some Changes among the San under the Influence of Relocation Plan in Botswana. In D. Anderson & K. Ikeya (Eds.), *Parks, Property, and Power: Managing Hunting Practice and Identity Within State Poicy Regimes* (pp. 183–198). Osaka: National Museum of Ethnology.

Iliev, I. D., Funari, V. A., Taylor, K. D., Nguyen, Q., Reyes, C. N., Strom, S. P., … Underhill, D. M. (2012). Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. *Science (New York, N.Y.)*, *336*(6086), 1314–1317. https://doi.org/10.1126/science.1221789

Ingram, C. J. E., Elamin, M. F., Mulcare, C. A., Weale, M. E., Tarekegn, A., Raga, T. O., … Swallow, D. M. (2007). A novel polymorphism associated with lactose

tolerance in Africa: Multiple causes for lactase persistence? *Human Genetics*, *120*(6), 779–788. https://doi.org/10.1007/s00439-006-0291-1

Inoguchi, S., Ohashi, Y., Narai-Kanayama, A., Aso, K., Nakagaki, T., & Fujisawa, T. (2012). Effects of non-fermented and fermented soybean milk intake on faecal microbiota and faecal metabolites in humans. *International Journal of Food Sciences and Nutrition*, *63*(4), 402–410. https://doi.org/10.3109/09637486.2011.630992

Isolauri, E. (2004). Dietary modification of atopic disease: Use of probiotics in the prevention of atopic dermatitis. *Current Allergy and Asthma Reports*, *4*(4), 270–275.

Issa, S. Y., Badran, E. F., Akl, K. F., & Shehabi, A. A. (2011). Epidemiological characteristics of Candida species colonizing oral and rectal sites of Jordanian infants. *BMC Pediatrics*, *11*(1), 79. https://doi.org/10.1186/1471-2431-11-79

Jackson, M. A., Jackson, M., Jeffery, I. B., Beaumont, M., Bell, J. T., Clark, A. G., … Steves, C. J. (2016). Signatures of early frailty in the gut microbiota. *Genome Medicine*, *8*(1), 8. https://doi.org/10.1186/s13073-016-0262-7

Jakobsson, H. E., Jernberg, C., Andersson, A. F., Sjölund-Karlsson, M., Jansson, J. K., & Engstrand, L. (2010). Short-Term Antibiotic Treatment Has Differing Long-Term Impacts on the Human Throat and Gut Microbiome. *PLoS ONE*, *5*(3), e9836. https://doi.org/10.1371/journal.pone.0009836

Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. R., VanLiere, J. M., Fung, H.-C., … Singleton, A. B. (2008). Genotype, haplotype and copy-number variation in

worldwide human populations. *Nature*, *451*(7181), 998–1003. https://doi.org/10.1038/nature06742

Jallow, M., Teo, Y. Y., Small, K. S., Rockett, K. A., Deloukas, P., Clark, T. G., … Kwiatkowski, D. P. (2009). Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nature Genetics*, *41*(6), 657–665. https://doi.org/10.1038/ng.388

Jantchou, P., Morois, S., Clavel-Chapelon, F., Boutron-Ruault, M.-C., & Carbonnel, F. (2010). Animal protein intake and risk of inflammatory bowel disease: The E3N prospective study. *The American Journal of Gastroenterology*, *105*(10), 2195–2201. https://doi.org/10.1038/ajg.2010.192

Jantermtor, S., Pinlaor, P., Sawadpanich, K., Pinlaor, S., Sangka, A., Wilailuckana, C., … Yoshikawa, H. (2013). Subtype identification of Blastocystis spp. Isolated from patients in a major hospital in northeastern Thailand. *Parasitology Research*, *112*(4), 1781–1786. https://doi.org/10.1007/s00436-012-3218-x

Jarvis, J. P., Scheinfeldt, L. B., Soi, S., Lambert, C., Omberg, L., Ferwerda, B., … Tishkoff, S. A. (2012). Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. *PLoS Genetics*, *8*(4), e1002641. https://doi.org/10.1371/journal.pgen.1002641

Jenkins, T. P., Formenti, F., Castro, C., Piubelli, C., Perandin, F., Buonfrate, D., … Cantacessi, C. (2018). A comprehensive analysis of the faecal microbiome and metabolome of Strongyloides stercoralis infected volunteers from a non-endemic area. *Scientific Reports*, *8*. https://doi.org/10.1038/s41598-018-33937-3

Jenkins, T. P., Rathnayaka, Y., Perera, P. K., Peachey, L. E., Nolan, M. J., Krause, L., … Cantacessi, C. (2017). Infections by human gastrointestinal helminths are associated with changes in faecal microbiota diversity and composition. *PLoS ONE*, *12*(9). https://doi.org/10.1371/journal.pone.0184719

Jernberg, C., Löfmark, S., Edlund, C., & Jansson, J. K. (2007). Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *The ISME Journal*, *1*(1), 56–66. https://doi.org/10.1038/ismej.2007.3

Jernberg, C., Löfmark, S., Edlund, C., & Jansson, J. K. (2010). Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology (Reading, England)*, *156*(Pt 11), 3216–3223. https://doi.org/10.1099/mic.0.040618-0

Jervis-Bardy, J., Leong, L. E. X., Marri, S., Smith, R. J., Choo, J. M., Smith-Vaughan, H. C., … Marsh, R. L. (2015). Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. *Microbiome*, *3*, 19. https://doi.org/10.1186/s40168-015-0083-8

Jiang, L., & Patel, S. N. (2008). Community assembly in the presence of disturbance: A microcosm experiment. *Ecology*, *89*(7), 1931–1940.

Jin, J.-S., Touyama, M., Hisada, T., & Benno, Y. (2012). Effects of green tea consumption on human fecal microbiota with special reference to Bifidobacterium species. *Microbiology and Immunology*, *56*(11), 729–739. https://doi.org/10.1111/j.1348-0421.2012.00502.x

Jin, U.-H., Lee, S.-O., Sridharan, G., Lee, K., Davidson, L. A., Jayaraman, A., … Safe, S. (2014). Microbiome-derived tryptophan metabolites and their aryl hydrocarbon

receptor-dependent agonist and antagonist activities. *Molecular Pharmacology*, *85*(5), 777–788. https://doi.org/10.1124/mol.113.091165

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2004). *2004 report on the global AIDS epidemic: 4th global report.* Retrieved from http://www.unaids.org/bankok2004/GAR2004_html/

Jothikumar, N., da Silva, A. J., Moura, I., Qvarnstrom, Y., & Hill, V. R. (2008). Detection and differentiation of Cryptosporidium hominis and Cryptosporidium parvum by dual TaqMan assays. *Journal of Medical Microbiology*, *57*(Pt 9), 1099–1105. https://doi.org/10.1099/jmm.0.2008/001461-0

Kadmon, R., & Benjamini, Y. (2006). Effects of Productivity and Disturbance on Species Richness: A Neutral Model. *The American Naturalist*, *167*(6), 939–946. https://doi.org/10.1086/504602

Kalliomäki, M., Salminen, S., Arvilommi, H., Kero, P., Koskinen, P., & Isolauri, E. (2001). Probiotics in primary prevention of atopic disease: A randomised placebo-controlled trial. *Lancet*, *357*(9262), 1076–1079. https://doi.org/10.1016/S0140-6736(00)04259-8

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kang, S., Denman, S. E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., & McSweeney, C. S. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflammatory Bowel Diseases*, *16*(12), 2034–2042. https://doi.org/10.1002/ibd.21319

Kao, D., Hotte, N., Gillevet, P., & Madsen, K. (2014). Fecal microbiota transplantation inducing remission in Crohn's colitis and the associated changes in fecal microbial profile. *Journal of Clinical Gastroenterology*, *48*(7), 625–628. https://doi.org/10.1097/MCG.0000000000000131

Kapiki, A., Costalos, C., Oikonomidou, C., Triantafyllidou, A., Loukatou, E., & Pertrohilou, V. (2007). The effect of a fructo-oligosaccharide supplemented formula on gut flora of preterm infants. *Early Human Development*, *83*(5), 335–339. https://doi.org/10.1016/j.earlhumdev.2006.07.003

Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L., & Gordon, J. I. (2011). Human nutrition, the gut microbiome and the immune system. *Nature*, *474*(7351), 327–336. https://doi.org/10.1038/nature10213

Kedia, G., Vázquez, J. A., Charalampopoulos, D., & Pandiella, S. S. (2009). In vitro fermentation of oat bran obtained by debranning with a mixed culture of human fecal bacteria. *Current Microbiology*, *58*(4), 338–342. https://doi.org/10.1007/s00284-008-9335-1

Keim, N. L., & Martin, R. J. (2014). Dietary whole grain–microbiota interactions: Insights into mechanisms for human health. *Advances in Nutrition (Bethesda, Md.)*, *5*(5), 556–557.

Kelly, P. (2010). Nutrition, intestinal defence and the microbiome. *The Proceedings of the Nutrition Society*, *69*(2), 261–268. https://doi.org/10.1017/S0029665110000108

Kettleson, E. M., Adhikari, A., Vesper, S., Coombs, K., Indugula, R., & Reponen, T. (2015). Key determinants of the fungal and bacterial microbiomes in homes.

Environmental Research, *138*, 130–135. https://doi.org/10.1016/j.envres.2015.02.003

Khoruts, A., Dicksved, J., Jansson, J. K., & Sadowsky, M. J. (2010). Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *Journal of Clinical Gastroenterology*, *44*(5), 354–360. https://doi.org/10.1097/MCG.0b013e3181c87e02

Kim, B. Y., & Lohmueller, K. E. (2015). Selection and reduced population size cannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. *American Journal of Human Genetics*, *96*(3), 454–461. https://doi.org/10.1016/j.ajhg.2014.12.029

Kim, C. H., Park, J., & Kim, M. (2014). Gut Microbiota-Derived Short-Chain Fatty Acids, T Cells, and Inflammation. *Immune Network*, *14*(6), 277–288. https://doi.org/10.4110/in.2014.14.6.277

Kim, D., Hofstaedter, C. E., Zhao, C., Mattei, L., Tanes, C., Clarke, E., … Bittinger, K. (2017). Optimizing methods and dodging pitfalls in microbiome research. *Microbiome*, *5*(1), 52. https://doi.org/10.1186/s40168-017-0267-5

Kim, M.-S., Hwang, S.-S., Park, E.-J., & Bae, J.-W. (2013). Strict vegetarian diet improves the risk factors associated with metabolic diseases by modulating gut microbiota and reducing intestinal inflammation. *Environmental Microbiology Reports*, *5*(5), 765–775. https://doi.org/10.1111/1758-2229.12079

Kirk, M. D., Pires, S. M., Black, R. E., Caipo, M., Crump, J. A., Devleesschauwer, B., … Angulo, F. J. (2015). World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral

Diseases, 2010: A Data Synthesis. *PLOS Medicine*, *12*(12), e1001921. https://doi.org/10.1371/journal.pmed.1001921

Ko, W.-Y., Kaercher, K. A., Giombini, E., Marcatili, P., Froment, A., Ibrahim, M., … Tishkoff, S. A. (2011). Effects of Natural Selection and Gene Conversion on the Evolution of Human Glycophorins Coding for MNS Blood Polymorphisms in Malaria-Endemic African Populations. *American Journal of Human Genetics*, *88*(6), 741–754. https://doi.org/10.1016/j.ajhg.2011.05.005

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., … Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences*, *108*(Supplement 1), 4578–4585. https://doi.org/10.1073/pnas.1000081107

Koren, O., Goodrich, J. K., Cullender, T. C., Spor, A., Laitinen, K., Kling Bäckhed, H., … Ley, R. E. (2012). Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy. *Cell*, *150*(3), 470–480. https://doi.org/10.1016/j.cell.2012.07.008

Koren, O., Knights, D., Gonzalez, A., Waldron, L., Segata, N., Knight, R., … Ley, R. E. (2013). A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. *PLoS Comput Biol*, *9*(1), e1002863. https://doi.org/10.1371/journal.pcbi.1002863

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform.

Applied and Environmental Microbiology, AEM.01043-13. https://doi.org/10.1128/AEM.01043-13

Kris-Etherton, P. M., Harris, W. S., Appel, L. J., & American Heart Association. Nutrition Committee. (2002). Fish consumption, fish oil, omega-3 fatty acids, and cardiovascular disease. *Circulation*, *106*(21), 2747–2757.

Labrie, V., Buske, O. J., Oh, E., Jeremian, R., Ptak, C., Gasiūnas, G., … Petronis, A. (2016). Lactase nonpersistence is directed by DNA-variation-dependent epigenetic aging. *Nature Structural & Molecular Biology*, *23*(6), 566–573. https://doi.org/10.1038/nsmb.3227

Lachance, J. (2010). Disease-associated alleles in genome-wide association studies are enriched for derived low frequency alleles relative to HapMap and neutral expectations. *BMC Medical Genomics*, *3*, 57. https://doi.org/10.1186/1755-8794-3-57

Lachance, J., Vernot, B., Elbers, C. C., Ferwerda, B., Froment, A., Bodo, J.-M., … Tishkoff, S. A. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell*, *150*(3), 457–469. https://doi.org/10.1016/j.cell.2012.07.009

Laforest-Lapointe, I., & Arrieta, M.-C. (2018). Microbial Eukaryotes: A Missing Link in Gut Microbiome Studies. *MSystems*, *3*(2). https://doi.org/10.1128/mSystems.00201-17

Lane, D. (1991). *16S/23S rRNA sequencing. In: Nucleic acid techniques in bacterial systematics. Stackebrandt, E., and Goodfellow, M., eds.* New York: John Wiley and Sons.

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., … Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, *31*(9), 814–821. https://doi.org/10.1038/nbt.2676

Lauder, A. P., Roche, A. M., Sherrill-Mix, S., Bailey, A., Laughlin, A. L., Bittinger, K., … Bushman, F. D. (2016). Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome*, *4*(1), 29. https://doi.org/10.1186/s40168-016-0172-3

Lawn, S. D., Butera, S. T., & Shinnick, T. M. (2002). Tuberculosis unleashed: The impact of human immunodeficiency virus infection on the host granulomatous response to Mycobacterium tuberculosis. *Microbes and Infection*, *4*(6), 635–646. https://doi.org/10.1016/S1286-4579(02)01582-4

Lay, C., Rigottier-Gois, L., Holmstrøm, K., Rajilic, M., Vaughan, E. E., de Vos, W. M., … Doré, J. (2005). Colonic microbiota signatures across five northern European countries. *Applied and Environmental Microbiology*, *71*(7), 4153–4155. https://doi.org/10.1128/AEM.71.7.4153-4155.2005

Lecomte, V., Kaakoush, N. O., Maloney, C. A., Raipuria, M., Huinao, K. D., Mitchell, H. M., & Morris, M. J. (2015). Changes in Gut Microbiota in Rats Fed a High Fat Diet Correlate with Obesity-Associated Metabolic Parameters. *PLoS ONE*, *10*(5). https://doi.org/10.1371/journal.pone.0126931

Lee, H. C., Jenner, A. M., Low, C. S., & Lee, Y. K. (2006). Effect of tea phenolics and their aromatic fecal bacterial metabolites on intestinal microbiota. *Research in Microbiology*, *157*(9), 876–884. https://doi.org/10.1016/j.resmic.2006.07.004

Lee, R. (1979). *The !Kung San. Men, Women, and Work in a Foraging Society.* Cambridge University Press.

Lee, S. C., Tang, M. S., Lim, Y. A. L., Choy, S. H., Kurtz, Z. D., Cox, L. M., … Loke, P. (2014). Helminth Colonization Is Associated with Increased Diversity of the Gut Microbiota. *PLoS Negl Trop Dis*, *8*(5), e2880. https://doi.org/10.1371/journal.pntd.0002880

Leffel, E., & Reed, D. (2004). Marburg and Ebola viruses as aerosol threats. - PubMedâ€ ”NCBI. *Biosecur Bioterror*, *2*(3), 186–191.

Leiby, J. S., McCormick, K., Sherrill-Mix, S., Clarke, E. L., Kessler, L. R., Taylor, L. J., … Bushman, F. D. (2018). Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. *Microbiome*, *6*(1), 196. https://doi.org/10.1186/s40168-018-0575-4

Leitch, E. C. M., Walker, A. W., Duncan, S. H., Holtrop, G., & Flint, H. J. (2007). Selective colonization of insoluble substrates by human faecal bacteria. *Environmental Microbiology*, *9*(3), 667–679. https://doi.org/10.1111/j.1462-2920.2006.01186.x

Leong, K. S. W., Derraik, J. G. B., Hofman, P. L., & Cutfield, W. S. (2018). Antibiotics, gut microbiome and obesity. *Clinical Endocrinology*, *88*(2), 185–200. https://doi.org/10.1111/cen.13495

Lepage, P., Colombet, J., Marteau, P., Sime-Ngando, T., Doré, J., & Leclerc, M. (2008). Dysbiosis in inflammatory bowel disease: A role for bacteriophages? *Gut*, *57*(3), 424–425. https://doi.org/10.1136/gut.2007.134668

Leung, J. M., Graham, A. L., & Knowles, S. C. L. (2018). Parasite-Microbiota Interactions With the Vertebrate Gut: Synthesis Through an Ecological Lens. *Frontiers in Microbiology, 9*. https://doi.org/10.3389/fmicb.2018.00843

Leung, J. M., & Loke, P. (2013). A role for IL-22 in the relationship between intestinal helminths, gut microbiota and mucosal immunity. *International Journal for Parasitology*, *43*(3–4), 253–257. https://doi.org/10.1016/j.ijpara.2012.10.015

Levine, M. E., Suarez, J. A., Brandhorst, S., Balasubramanian, P., Cheng, C.-W., Madia, F., … Longo, V. D. (2014). Low protein intake is associated with a major reduction in IGF-1, cancer, and overall mortality in the 65 and younger but not older population. *Cell Metabolism*, *19*(3), 407–417. https://doi.org/10.1016/j.cmet.2014.02.006

Lewis, S., & Cochrane, S. (2007). Alteration of sulfate and hydrogen metabolism in the human colon by changing intestinal transit rate. *The American Journal of Gastroenterology*, *102*(3), 624–633. https://doi.org/10.1111/j.1572-0241.2006.01020.x

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 11070–11075. https://doi.org/10.1073/pnas.0504978102

Ley, R. E., Peterson, D. A., & Gordon, J. I. (2006). Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. *Cell*, *124*(4), 837–848. https://doi.org/10.1016/j.cell.2006.02.017

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Microbial ecology: Human gut microbes associated with obesity. *Nature*, *444*(7122), 1022–1023. https://doi.org/10.1038/4441022a

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., … Zhao, L. (2008). Symbiotic gut microbes modulate human metabolic phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(6), 2117–2122. https://doi.org/10.1073/pnas.0712038105

Li, S., Schlebusch, C., & Jakobsson, M. (2014). Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1793), 20141448. https://doi.org/10.1098/rspb.2014.1448

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22.

Lin, A., Bik, E. M., Costello, E. K., Dethlefsen, L., Haque, R., Relman, D. A., & Singh, U. (2013). Distinct Distal Gut Microbiome Diversity and Composition in Healthy Children from Bangladesh and the United States. *PLOS ONE*, *8*(1), e53838. https://doi.org/10.1371/journal.pone.0053838

Link-Amster, H., Rochat, F., Saudan, K. Y., Mignot, O., & Aeschlimann, J. M. (1994). Modulation of a specific humoral immune response and changes in intestinal flora

mediated through fermented milk intake. *FEMS Immunology and Medical Microbiology*, *10*(1), 55–63.

Littman, D. R., & Pamer, E. G. (2011). Role of the commensal microbiota in normal and pathogenic host immune responses. *Cell Host & Microbe*, *10*(4), 311–323. https://doi.org/10.1016/j.chom.2011.10.004

Liu, J.-E., Zhang, Y., Zhang, J., Dong, P.-L., Chen, M., & Duan, Z.-P. (2010). Probiotic yogurt effects on intestinal flora of patients with chronic liver disease. *Nursing Research*, *59*(6), 426–432. https://doi.org/10.1097/NNR.0b013e3181fa4dc6

Liu, Z., Lin, X., Huang, G., Zhang, W., Rao, P., & Ni, L. (2014). Prebiotic effects of almonds and almond skins on intestinal microbiota in healthy adult humans. *Anaerobe*, *26*, 1–6. https://doi.org/10.1016/j.anaerobe.2013.11.007

Llorente, M. G., Jones, E. R., Eriksson, A., Siska, V., Arthur, K. W., Arthur, J. W., … Manica, A. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science*, *350*(6262), 820–822. https://doi.org/10.1126/science.aad2879

Lockyer, S., & Nugent, A. P. (2017). Health effects of resistant starch. *Nutrition Bulletin*, *42*(1), 10–41. https://doi.org/10.1111/nbu.12244

Lokki, A. I., Järvelä, I., Israelsson, E., Maiga, B., Troye-Blomberg, M., Dolo, A., … Holmberg, V. (2011). Lactase persistence genotypes and malaria susceptibility in Fulani of Mali. *Malaria Journal*, *10*(1), 9. https://doi.org/10.1186/1475-2875-10-9

Lopez-Legarrea, P., Fuller, N. R., Zulet, M. A., Martinez, J. A., & Caterson, I. D. (2014). The influence of Mediterranean, carbohydrate and high protein diets on gut

microbiota composition in the treatment of obesity and associated inflammatory state. *Asia Pacific Journal of Clinical Nutrition*, *23*(3), 360–368. https://doi.org/10.6133/apjcn.2014.23.3.16

Lorenzo Pisarello, M. J., Vintiñi, E. O., González, S. N., Pagani, F., & Medina, M. S. (2015). Decrease in lactobacilli in the intestinal microbiota of celiac children with a gluten-free diet, and selection of potentially probiotic strains. *Canadian Journal of Microbiology*, *61*(1), 32–37. https://doi.org/10.1139/cjm-2014-0472

Losasso, C., Eckert, E. M., Mastrorilli, E., Villiger, J., Mancin, M., Patuzzi, I., … Ricci, A. (2018). Assessing the Influence of Vegan, Vegetarian and Omnivore Oriented Westernized Dietary Styles on Human Gut Microbiota: A Cross Sectional Study. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.00317

Lozupone, C., & Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, *71*(12), 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005

Lozupone, C.A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, *489*(7415), 220–230. https://doi.org/10.1038/nature11550

Lozupone, Catherine A., Hamady, M., Kelley, S. T., & Knight, R. (2007). Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Applied and Environmental Microbiology*, *73*(5), 1576–1585. https://doi.org/10.1128/AEM.01996-06

Lozupone, Catherine A., Li, M., Campbell, T. B., Flores, S. C., Linderman, D., Gebert, M. J., … Palmer, B. E. (2013). Alterations in the gut microbiota associated with

HIV-1 infection. *Cell Host & Microbe*, *14*(3). https://doi.org/10.1016/j.chom.2013.08.006

Lozupone, Catherine A., Rhodes, M. E., Neff, C. P., Fontenot, A. P., Campbell, T. B., & Palmer, B. E. (2014). HIV-induced alteration in gut microbiota: Driving factors, consequences, and effects of antiretroviral therapy. *Gut Microbes*, *5*(4), 562–570. https://doi.org/10.4161/gmic.32132

Machiels, K., Joossens, M., Sabino, J., Preter, V. D., Arijs, I., Eeckhaut, V., … Vermeire, S. (2014). A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut*, *63*(8), 1275–1283. https://doi.org/10.1136/gutjnl-2013-304833

Macholdt, E., Slatkin, M., Pakendorf, B., & Stoneking, M. (2015). New insights into the history of the C-14010 lactase persistence variant in Eastern and Southern Africa. *Am J Phys Anthropol*, *156*(4), 661–664. https://doi.org/10.1016/bs.mcb.2015.01.016.Observing

MacNeil, A., & Rollin, P. E. (2012). Ebola and Marburg Hemorrhagic Fevers: Neglected Tropical Diseases? *PLOS Neglected Tropical Diseases*, *6*(6), e1546. https://doi.org/10.1371/journal.pntd.0001546

Macpherson, A. J., & Harris, N. L. (2004). Interactions between commensal intestinal bacteria and the immune system. *Nature Reviews. Immunology*, *4*(6), 478–485. https://doi.org/10.1038/nri1373

Mafessoni, F. (2019). Encounters with archaic hominins. *Nature Ecology & Evolution*, *3*(1), 14. https://doi.org/10.1038/s41559-018-0729-6

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ, 2*. https://doi.org/10.7717/peerj.593

Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M. R., Hallows, J. L., … Zink, A. (2016). The 5300-year-old Helicobacter pylori genome of the Iceman. *Science (New York, N.Y.), 351*(6269), 162–165. https://doi.org/10.1126/science.aad2545

Mäkivuokko, H., Tiihonen, K., Tynkkynen, S., Paulin, L., & Rautonen, N. (2010). The effect of age and non-steroidal anti-inflammatory drugs on human intestinal microbiota composition. *The British Journal of Nutrition, 103*(2), 227–234. https://doi.org/10.1017/S0007114509991553

Malmuthuge, N., Griebel, P. J., & Guan, L. L. (2014). Taxonomic identification of commensal bacteria associated with the mucosa and digesta throughout the gastrointestinal tracts of preweaned calves. *Applied and Environmental Microbiology*, *80*(6), 2021–2028. https://doi.org/10.1128/AEM.03864-13

Mandal, S., Treuren, W. V., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: A novel method for studying microbial composition. *Microbial Ecology in Health and Disease, 26*(0). https://doi.org/10.3402/mehd.v26.27663

Marchesi, J. R. (2014). *The Human Microbiota and Microbiome*. CABI.

Marcobal, A., Barboza, M., Froehlich, J. W., Block, D. E., German, J. B., Lebrilla, C. B., & Mills, D. A. (2010). Consumption of human milk oligosaccharides by gut-

related microbes. *Journal of Agricultural and Food Chemistry*, *58*(9), 5334–5340. https://doi.org/10.1021/jf9044205

Marlowe, F. (2010). *The Hadza: Hunter-Gatherers of Tanzania*. University of California Press.

Marlowe, F., & Berbesque, C. (n.d.). Tubers as fallback foods and their impact on Hadza hunter-gatherers. *American Journal of Physical Anthropology*, *140*, 751–758.

Marshall, B. J., Armstrong, J. A., McGechie, D. B., & Glancy, R. J. (1985). Attempt to fulfil Koch's postulates for pyloric Campylobacter. *The Medical Journal of Australia*, *142*(8), 436–439.

Martin, F.-P. J., Montoliu, I., Nagy, K., Moco, S., Collino, S., Guy, P., … Kochhar, S. (2012). Specific Dietary Preferences Are Linked to Differing Gut Microbial Metabolic Activity in Response to Dark Chocolate Intake. *Journal of Proteome Research*, *11*(12), 6252–6263. https://doi.org/10.1021/pr300915z

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Martínez, I., Lattimer, J. M., Hubach, K. L., Case, J. A., Yang, J., Weber, C. G., … Walter, J. (2013). Gut microbiome composition is linked to whole grain-induced immunological improvements. *The ISME Journal*, *7*(2), 269–280. https://doi.org/10.1038/ismej.2012.104

Martínez, I., Stegen, J. C., Maldonado-Gómez, M. X., Eren, A. M., Siba, P. M., Greenhill, A. R., & Walter, J. (2015). The Gut Microbiota of Rural Papua New

Guineans: Composition, Diversity Patterns, and Ecological Processes. *Cell Reports*, *11*(4), 527–538. https://doi.org/10.1016/j.celrep.2015.03.049

Martinez, K. B., Leone, V., & Chang, E. B. (2017). Western diets, gut dysbiosis, and metabolic diseases: Are they linked? *Gut Microbes*, *8*(2), 130–142. https://doi.org/10.1080/19490976.2016.1270811

Maslowski, K. M., Vieira, A. T., Ng, A., Kranich, J., Sierro, F., Yu, D., … Mackay, C. R. (2009). Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature*, *461*(7268), 1282–1286. https://doi.org/10.1038/nature08530

Matarazzo, F., Ribeiro, A. C., Faveri, M., Taddei, C., Martinez, M. B., & Mayer, M. P. A. (2012). The domain Archaea in human mucosal surfaces. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *18*(9), 834–840. https://doi.org/10.1111/j.1469-0691.2012.03958.x

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., … Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499–503. https://doi.org/10.1038/nature16152

Matsumoto, K., Takada, T., Shimizu, K., Moriyama, K., Kawakami, K., Hirano, K., … Nomoto, K. (2010). Effects of a probiotic fermented milk beverage containing Lactobacillus casei strain Shirota on defecation frequency, intestinal microbiota, and the intestinal environment of healthy individuals with soft stools. *Journal of Bioscience and Bioengineering*, *110*(5), 547–552. https://doi.org/10.1016/j.jbiosc.2010.05.016

McDonald, D, Price, M., Goodrich, J., Nawrocki, E., DeSantis, T., Probst, A., … Hegenholtz, P. (2012). An Improved Greengenes taxonomy with explicity ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J*, *6*(3), 610–618.

McDonald, Daniel, Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., … Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, *6*(3), 610–618. https://doi.org/10.1038/ismej.2011.139

McDougall, I., Brown, F. H., & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature*, *433*(7027), 733–736. https://doi.org/10.1038/nature03258

McHardy, I. H., Li, X., Tong, M., Ruegger, P., Jacobs, J., Borneman, J., … Braun, J. (2013). HIV Infection is associated with compositional and functional shifts in the rectal mucosal microbiota. *Microbiome*, *1*(1), 26. https://doi.org/10.1186/2049-2618-1-26

McKenna, P., Hoffmann, C., Minkah, N., Aye, P. P., Lackner, A., Liu, Z., … Bushman, F. D. (2008). The Macaque Gut Microbiome in Health, Lentiviral Infection, and Chronic Enterocolitis. *PLoS Pathog*, *4*(2), e20. https://doi.org/10.1371/journal.ppat.0040020

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Meddah, A. T., Yazourh, A., Desmet, I., Risbourg, B., Verstraete, W., & Romond, M. B. (2001). The regulatory effects of whey retentate from bifidobacteria fermented milk on the microbiota of the Simulator of the Human Intestinal Microbial Ecosystem (SHIME). *Journal of Applied Microbiology*, *91*(6), 1110–1117.

Meisel, J. S., Hannigan, G. D., Tyldsley, A. S., SanMiguel, A. J., Hodkinson, B. P., Zheng, Q., & Grice, E. A. (2016). Skin microbiome surveys are strongly influenced by experimental design. *The Journal of Investigative Dermatology*, *136*(5), 947–956. https://doi.org/10.1016/j.jid.2016.01.016

Mejia, R., Vicuña, Y., Broncano, N., Sandoval, C., Vaca, M., Chico, M., … Nutman, T. B. (2013). A novel, multi-parallel, real-time polymerase chain reaction approach for eight gastrointestinal parasites provides improved diagnostic capabilities to resource-limited at-risk populations. *The American Journal of Tropical Medicine and Hygiene*, *88*(6), 1041–1047. https://doi.org/10.4269/ajtmh.12-0726

Meyer, A. M., Ramzan, N. N., Loftus, E. V., Heigh, R. I., & Leighton, J. A. (2004). The diagnostic yield of stool pathogen studies during relapses of inflammatory bowel disease. *Journal of Clinical Gastroenterology*, *38*(9), 772–775.

Migliano, A. B., Romero, I. G., Leavesley, M., & Pagani, L. (2013). *Evolution of the Pygmy Phenotype: Evidence of Positive Selection from Genome-wide Scans in African , Asian , and Melanesian Pygmies Evolution of the Pygmy Phenotype: Evidence of Positive Selection from*. *85*(1).

Mihajlovski, A., Doré, J., Levenez, F., Alric, M., & Brugère, J.-F. (2010). Molecular evaluation of the human gut methanogenic archaeal microbiota reveals an age-

associated increase of the diversity. *Environmental Microbiology Reports*, *2*(2), 272–280. https://doi.org/10.1111/j.1758-2229.2009.00116.x

Million, M., Maraninchi, M., Henry, M., Armougom, F., Richet, H., Carrieri, P., … Raoult, D. (2012). Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii. *International Journal of Obesity (2005)*, *36*(6), 817–825. https://doi.org/10.1038/ijo.2011.153

Minot, S., Bryson, A., Chehoud, C., Wu, G. D., Lewis, J. D., & Bushman, F. D. (2013). Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(30), 12450–12455. https://doi.org/10.1073/pnas.1300833110

Mira, A., Pushker, R., & Rodríguez-Valera, F. (2006). The Neolithic revolution of bacterial genomes. *Trends in Microbiology*, *14*(5), 200–206. https://doi.org/10.1016/j.tim.2006.03.001

Moeller, A. H., Li, Y., Ngole, E. M., Ahuka-Mundeke, S., Lonsdorf, E. V., Pusey, A. E., … Ochman, H. (2014). Rapid changes in the gut microbiome during human evolution. *Proceedings of the National Academy of Sciences*, 201419136. https://doi.org/10.1073/pnas.1419136111

Monaco, C. L., Gootenberg, D. B., Zhao, G., Handley, S. A., Ghebremichael, M. S., Lim, E. S., … Virgin, H. W. (2016). Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host & Microbe*, *19*(3), 311–322. https://doi.org/10.1016/j.chom.2016.02.011

Montinaro, F., Busby, G. B. J., Gonzalez-santos, M., & Oosthuitzen, O. (2016). Complex ancient genetic structure and cultural transitions in southern African populations. *BioRxiv*. https://doi.org/10.1101/043562

Morgan, X. C., & Huttenhower, C. (2012). Chapter 12: Human Microbiome Analysis. *PLoS Comput Biol*, *8*(12), e1002808. https://doi.org/10.1371/journal.pcbi.1002808

Morris, A., & Nicholson, G. (1987). Ingestion of Campylobacter pyloridis causes gastritis and raised fasting gastric pH. *The American Journal of Gastroenterology*, *82*(3), 192–199.

Morton, E. R., Lynch, J., Froment, A., Lafosse, S., Heyer, E., Przeworski, M., … Ségurel, L. (2015). Variation in Rural African Gut Microbiota Is Strongly Correlated with Colonization by Entamoeba and Subsistence. *PLOS Genet*, *11*(11), e1005658. https://doi.org/10.1371/journal.pgen.1005658

Mudd, J. C., & Brenchley, J. M. (2016). Gut Mucosal Barrier Dysfunction, Microbial Dysbiosis, and Their Role in HIV-1 Disease Progression. *The Journal of Infectious Diseases*, *214 Suppl 2*, S58-66. https://doi.org/10.1093/infdis/jiw258

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., González, A., Fontana, L., … Gordon, J. I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science (New York, N.Y.)*, *332*(6032), 970–974. https://doi.org/10.1126/science.1198719

Mueller, N. T., Bakacs, E., Combellick, J., Grigoryan, Z., & Dominguez-Bello, M. G. (2015). The infant microbiome development: Mom matters. *Trends in Molecular Medicine*, *21*(2), 109–117. https://doi.org/10.1016/j.molmed.2014.12.002

Mueller, S., Saunier, K., Hanisch, C., Norin, E., Alm, L., Midtvedt, T., … Blaut, M. (2006). Differences in fecal microbiota in different European study populations in relation to age, gender, and country: A cross-sectional study. *Applied and Environmental Microbiology*, *72*(2), 1027–1033. https://doi.org/10.1128/AEM.72.2.1027-1033.2006

Mutlu, E. A., Keshavarzian, A., Losurdo, J., Swanson, G., Siewe, B., Forsyth, C., … Landay, A. (2014). A compositional look at the human gastrointestinal microbiome and immune activation parameters in HIV infected subjects. *PLoS Pathogens*, *10*(2), e1003829. https://doi.org/10.1371/journal.ppat.1003829

Myers, S. S., Gaffikin, L., Golden, C. D., Ostfeld, R. S., Redford, K. H., Ricketts, T. H., … Osofsky, S. A. (2013). Human health impacts of ecosystem alteration. *Proceedings of the National Academy of Sciences*, *110*(47), 18753–18760. https://doi.org/10.1073/pnas.1218656110

Nadal, I., Donat, E., Donant, E., Ribes-Koninckx, C., Calabuig, M., & Sanz, Y. (2007). Imbalance in the composition of the duodenal microbiota of children with coeliac disease. *Journal of Medical Microbiology*, *56*(Pt 12), 1669–1674. https://doi.org/10.1099/jmm.0.47410-0

Nestel, P. (1985). *Nutrition of Maasai women and children in relation to subsistence food production*. University of London, London, UK: Queen Elizabeth College.

Newman, J. L. (1970). *The Ecological Basis for Subsistence Change among the Sandawe of Tanzania.* Washington D.C.: National Academy of Sciences.

Ni, J., Shen, T.-C. D., Chen, E. Z., Bittinger, K., Bailey, A., Roggiani, M., … Wu, G. D. (2017). A role for bacterial urease in gut dysbiosis and Crohn's disease. *Science*

Translational Medicine, *9*(416), eaah6888. https://doi.org/10.1126/scitranslmed.aah6888

Nicolucci, A. C., Hume, M. P., Martínez, I., Mayengbam, S., Walter, J., & Reimer, R. A. (2017). Prebiotics Reduce Body Fat and Alter Intestinal Microbiota in Children Who Are Overweight or With Obesity. *Gastroenterology*, *153*(3), 711–722. https://doi.org/10.1053/j.gastro.2017.05.055

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, *541*(7637), 302–310. https://doi.org/10.1038/nature21347

Noverr, M. C., & Huffnagle, G. B. (2005). The "microflora hypothesis" of allergic diseases. *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology*, *35*(12), 1511–1520. https://doi.org/10.1111/j.1365-2222.2005.02379.x

Nowak, R. G., Bentzen, S. M., Ravel, J., Crowell, T. A., Dauda, W., Ma, B., … TRUSTRV368 Study Group. (2017). Rectal microbiota among HIV-uninfected, untreated HIV, and treated HIV-infected in Nigeria. *AIDS (London, England)*, *31*(6), 857–862. https://doi.org/10.1097/QAD.0000000000001409

Nuriel-Ohayon, M., Neuman, H., & Koren, O. (2016). Microbial Changes during Pregnancy, Birth, and Infancy. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.01031

Obregon-Tito, A. J., Tito, R. Y., Metcalf, J., Sankaranarayanan, K., Clemente, J. C., Ursell, L. K., … Lewis, C. M. (2015). Subsistence strategies in traditional

societies distinguish gut microbiomes. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms7505

Odds, F. C., Davidson, A. D., Jacobsen, M. D., Tavanti, A., Whyte, J. A., Kibbler, C. C., … Gow, N. a. R. (2006). Candida albicans strain maintenance, replacement, and microvariation demonstrated by multilocus sequence typing. *Journal of Clinical Microbiology*, *44*(10), 3647–3658. https://doi.org/10.1128/JCM.00934-06

Ogilvie, L. A., Bowler, L. D., Caplin, J., Dedi, C., Diston, D., Cheek, E., … Jones, B. V. (2013). Genome signature-based dissection of human gut metagenomes to extract subliminal viral sequences. *Nature Communications*, *4*, 2420. https://doi.org/10.1038/ncomms3420

O'Hara, A. M., & Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Reports*, *7*(7), 688–693. https://doi.org/10.1038/sj.embor.7400731

Ohenjo, N., Willis, R., Jackson, D., Nettleton, C., Good, K., & Mugarura, B. (2006). Health of Indigenous people in Africa. *The Lancet*, *367*(9526), 1937–1946. https://doi.org/10.1016/S0140-6736(06)68849-1

Okada, H., Kuhn, C., Feillet, H., & Bach, J.-F. (2010). The "hygiene hypothesis" for autoimmune and allergic diseases: An update. *Clinical and Experimental Immunology*, *160*(1), 1–9. https://doi.org/10.1111/j.1365-2249.2010.04139.x

Okansen, J. (2013). *Multivariate Analysis of Ecological Communities in R: vegan tutorial*.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., … Wagner, H. (2013). vegan: Community Ecology Package (Version 2.0-10). Retrieved from http://cran.r-project.org/web/packages/vegan/index.html

Org, E., Mehrabian, M., Parks, B. W., Shipkova, P., Liu, X., Drake, T. A., & Lusis, A. J. (2016). Sex differences and hormonal effects on gut microbiota composition in mice. *Gut Microbes*, *7*(4), 313–322. https://doi.org/10.1080/19490976.2016.1203502

Organization of African Unity. (2000). *Abuja declaration on HIV/AIDS, tuberculosis and other related infectious diseases*. Addis Ababa.

Osborne, L. C., Monticelli, L. A., Nice, T. J., Sutherland, T. E., Siracusa, M. C., Hepworth, M. R., … Artis, D. (2014). Virus-helminth coinfection reveals a microbiota-independent mechanism of immunomodulation. *Science*, *345*(6196), 578–582. https://doi.org/10.1126/science.1256942

Panda, S., El khader, I., Casellas, F., López Vivancos, J., García Cors, M., Santiago, A., … Manichanh, C. (2014). Short-term effect of antibiotics on human gut microbiota. *PloS One*, *9*(4), e95476. https://doi.org/10.1371/journal.pone.0095476

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analysis of Phylogenetics and Evolution in R Language. *Bioinformatics*, *20*(2), 289–290.

Parvin, S., Easmin, D., Sheikh, A., Biswas, M., Sharma, S. C. D., Jahan, M. G. S., … Shovon, M. S. (2015). Nutritional Analysis of Date Fruits (Phoenix dactylifera L.) in Perspective of Bangladesh. *American Journal of Life Sciences*, *3*(4), 274. https://doi.org/10.11648/j.ajls.20150304.14

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., … Segata, N. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*, *176*(3), 649-662.e20. https://doi.org/10.1016/j.cell.2019.01.001

Pasquereau, S., Kumar, A., & Herbein, G. (2017). Targeting TNF and TNF Receptor Pathway in HIV-1 Infection: From Immune Activation to Viral Reservoirs. *Viruses*, *9*(4). https://doi.org/10.3390/v9040064

Patin, E., Laval, G., Barreiro, L. B., Salas, A., Semino, O., Santachiara-Benerecetti, S., … Quintana-Murci, L. (2009). Inferring the Demographic History of African Farmers and Pygmy Hunter–Gatherers Using a Multilocus Resequencing Data Set. *PLOS Genet*, *5*(4), e1000448. https://doi.org/10.1371/journal.pgen.1000448

Paweska, J. T., Vuren, P. J. van, Masumu, J., Leman, P. A., Grobbelaar, A. A., Birkhead, M., … Kemp, A. (2012). Virological and Serological Findings in Rousettus aegyptiacus Experimentally Inoculated with Vero Cells-Adapted Hogan Strain of Marburg Virus. *PLOS ONE*, *7*(9), e45479. https://doi.org/10.1371/journal.pone.0045479

Payne, M. S., & Bayatibojakhi, S. (2014). Exploring preterm birth as a polymicrobial disease: An overview of the uterine microbiome. *Frontiers in Immunology*, *5*, 595. https://doi.org/10.3389/fimmu.2014.00595

Peabody, M. A., Van Rossum, T., Lo, R., & Brinkman, F. S. L. (2015). Evaluation of shotgun metagenomics sequence classification methods using in silico and in vitro simulated communities. *BMC Bioinformatics*, *16*(1), 362. https://doi.org/10.1186/s12859-015-0788-5

Pedro Soares, Rito, T., Pereira, L., & Richards, M. B. (2016). A genetic perspective on African prehistory. In *Africa from MIS 6-2: Population Dynamics and Paleoenvironments, Vertebrate* (pp. 383–405).

Penders, J., Thijs, C., van den Brandt, P. A., Kummeling, I., Snijders, B., Stelma, F., … Stobberingh, E. E. (2007). Gut microbiota composition and development of atopic manifestations in infancy: The KOALA Birth Cohort Study. *Gut*, *56*(5), 661–667. https://doi.org/10.1136/gut.2006.100164

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., … Stobberingh, E. E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*, *118*(2), 511–521. https://doi.org/10.1542/peds.2005-2824

Pennington, R. L. (2002). Economic stratification and health among the Herero of Botswana. In W. R. Leonard & M. H. Crawford (Eds.), *The Human Biology of Pastoral Populations*. Cambridge University Press.

Pérez, P. F., Minnaard, J., Rouvet, M., Knabenhans, C., Brassart, D., De Antoni, G. L., & Schiffrin, E. J. (2001). Inhibition of Giardia intestinalis by Extracellular Factors from Lactobacilli: An In Vitro Study. *Applied and Environmental Microbiology*, *67*(11), 5037–5042. https://doi.org/10.1128/AEM.67.11.5037-5042.2001

Perez-Muñoz, M. E., Arrieta, M.-C., Ramer-Tait, A. E., & Walter, J. (2017). A critical assessment of the “sterile womb” and “in utero colonization” hypotheses: Implications for research on the pioneer infant microbiome. *Microbiome*, *5*(1), 48. https://doi.org/10.1186/s40168-017-0268-4

Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., … Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*, *39*(10), 1256–1260. https://doi.org/10.1038/ng2123

Perry, G. H., Foll, M., Grenier, J.-C., Patin, E., Nédélec, Y., Pacis, A., … Barreiro, L. B. (2014). Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. *Proceedings of the National Academy of Sciences, 111*(35), E3596–E3603. https://doi.org/10.1073/pnas.1402875111

Peterfreund, G. L., Vandivier, L. E., Sinha, R., Marozsan, A. J., Olson, W. C., Zhu, J., & Bushman, F. D. (2012). Succession in the gut microbiome following antibiotic and antibody therapies for Clostridium difficile. *PloS One, 7*(10), e46966. https://doi.org/10.1371/journal.pone.0046966

Pickrell, J. K., Coop, G., Novembre, J., Kudaravalli, S., Li, J. Z., Absher, D., … Pritchard, J. K. (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Research, 19*(5), 826–837. https://doi.org/10.1101/gr.087577.108

Pickrell, J. K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., … Pakendorf, B. (2012). The genetic prehistory of southern Africa. *Nature Communications*, *3*, 1143. https://doi.org/10.1038/ncomms2140

Pickrell, J. K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., … Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(7), 2632–2637. https://doi.org/10.1073/pnas.1313787111

Pierson, D. L., Mehta, S. K., Magee, B. B., & Mishra, S. K. (1995). Person-to-person transfer of Candida albicans in the spacecraft environment. *Journal of Medical and Veterinary Mycology: Bi-Monthly Publication of the International Society for Human and Animal Mycology*, *33*(3), 145–150.

Plagnol, V., & Wall, J. D. (2006). Possible ancestral structure in human populations. *PLoS Genetics*, *2*(7), e105. https://doi.org/10.1371/journal.pgen.0020105

Plovier, H., Everard, A., Druart, C., Depommier, C., Van Hul, M., Geurts, L., … Cani, P. D. (2017). A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nature Medicine*, *23*(1), 107–113. https://doi.org/10.1038/nm.4236

Pontzer, H., Raichlen, D. A., Wood, B. M., Mabulla, A. Z. P., Racette, S. B., & Marlowe, F. W. (2012). Hunter-gatherer energetics and human obesity. *PloS One*, *7*(7), e40503. https://doi.org/10.1371/journal.pone.0040503

Pramono, A. K., Kuwahara, H., Itoh, T., Toyoda, A., Yamada, A., & Hongoh, Y. (2017). Discovery and Complete Genome Sequence of a Bacteriophage from an Obligate Intracellular Symbiont of a Cellulolytic Protist in the Termite Gut. *Microbes and Environments*, *32*(2), 112–117. https://doi.org/10.1264/jsme2.ME16175

Prendergast, M. E., Lipson, M., Sawchuk, E. A., Olalde, I., Ogola, C. A., Rohland, N., … Reich, D. (2019). Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science*, eaaw6275. https://doi.org/10.1126/science.aaw6275

Price, M., Dehal, P., & Arkin, A. (2010). FastTree2â€ ”Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*.

Prince, A. L., Antony, K. M., Chu, D. M., & Aagaard, K. M. (2014). The microbiome, parturition, and timing of birth: More questions than answers. *Journal of Reproductive Immunology*, *104–105*, 12–19. https://doi.org/10.1016/j.jri.2014.03.006

Qi, Q. G., Hu, T., & Zhou, X. D. (2005). Frequency, species and molecular characterization of oral Candida in hosts of different age in China. *Journal of Oral Pathology & Medicine*, *34*(6), 352–356. https://doi.org/10.1111/j.1600-0714.2005.00326.x

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., … Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. https://doi.org/10.1038/nature08821

Queipo-Ortuño, M. I., Boto-Ordóñez, M., Murri, M., Gomez-Zumaquero, J. M., Clemente-Postigo, M., Estruch, R., … Tinahones, F. J. (2012). Influence of red wine polyphenols and ethanol on the gut microbiota ecology and biochemical biomarkers. *The American Journal of Clinical Nutrition*, *95*(6), 1323–1334. https://doi.org/10.3945/ajcn.111.027847

Rajkumar, H., Mahmood, N., Kumar, M., Varikuti, S. R., Challa, H. R., & Myakala, S. P. (2014). Effect of Probiotic (VSL#3) and Omega-3 on Lipid Profile, Insulin Sensitivity, Inflammatory Markers, and Gut Colonization in Overweight Adults: A Randomized, Controlled Trial [Research article]. https://doi.org/10.1155/2014/348959

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(44), 15942–15947. https://doi.org/10.1073/pnas.0507611102

Ramanan, D., Bowcutt, R., Lee, S. C., Tang, M. S., Kurtz, Z. D., Ding, Y., … Cadwell, K. (2016). Helminth infection promotes colonization resistance via type 2 immunity. *Science (New York, N.Y.)*. https://doi.org/10.1126/science.aaf3229

Rampelli, S., Schnorr, S. L., Consolandi, C., Turroni, S., Severgnini, M., Peano, C., … Candela, M. (2015). Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. *Current Biology: CB*, *25*(13), 1682–1693. https://doi.org/10.1016/j.cub.2015.04.055

Ranciaro, A., Campbell, M. C., Hirbo, J. B., Ko, W.-Y., Froment, A., Anagnostou, P., … Tishkoff, S. A. (2014). Genetic Origins of Lactase Persistence and the Spread of Pastoralism in Africa. *The American Journal of Human Genetics*, *94*(4), 496–510. https://doi.org/10.1016/j.ajhg.2014.02.009

Rautava, S., Collado, M. C., Salminen, S., & Isolauri, E. (2012). Probiotics Modulate Host-Microbe Interaction in the Placenta and Fetal Gut: A Randomized, Double-Blind, Placebo-Controlled Trial. *Neonatology*, *102*(3), 178–184. https://doi.org/10.1159/000339182

Reddy, B. S., Weisburger, J. H., & Wynder, E. L. (1975). Effects of high risk and low risk diets for colon carcinogenesis on fecal microflora and steroids in man. *The Journal of Nutrition*, *105*(7), 878–884. https://doi.org/10.1093/jn/105.7.878

Reeves-Daniel, A. M., Iskandar, S. S., Bowden, D. W., Bostrom, M. A., Hicks, P. J., Comeau, M. E., … Freedman, B. I. (2010). Is collapsing C1q nephropathy another MYH9-associated kidney disease? A case report. *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation*, *55*(5), e21-24. https://doi.org/10.1053/j.ajkd.2009.10.060

Regional Office for Africa, Africa Health Organization, World Health Organization. (2016a). *Botswana: Factsheets of Health Statistics 2016.* Retrieved from http://www.aho.afro.who.int/profiles_information/images/0/05/Botswana-Statistical_Factsheet.pdf

Regional Office for Africa, Africa Health Organization, World Health Organization. (2016b). *United Republic of Tanzania: Factsheets of Health Statistics 2016*. Retrieved from http://www.aho.afro.who.int/profiles_information/images/c/c8/Tanzania-Statistical_Factsheet.pdf

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., … Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, *468*(7327), 1053–1060. https://doi.org/10.1038/nature09710

Renz, H., Holt, P. G., Inouye, M., Logan, A. C., Prescott, S. L., & Sly, P. D. (2017). An exposome perspective: Early-life events and immune development in a changing world. *Journal of Allergy and Clinical Immunology*, *140*(1), 24–40. https://doi.org/10.1016/j.jaci.2017.05.015

Reyes-Centeno, H. (2016). Out of Africa and into Asia: Fossil and genetic evidence on modern human origins and dispersals. *Quaternary International*. https://doi.org/10.1016/j.quaint.2015.11.063

Richerson, P. J., Boyd, R., & Henrich, J. (2010). Gene-culture coevolution in the age of genomics. *Proceedings of the National Academy of Sciences*, *107*(Supplement_2), 8985–8992. https://doi.org/10.1073/pnas.0914631107

Rightmire, G. P. (2009). Out of Africa: Modern human origins special feature: middle and later Pleistocene hominins in Africa and Southwest Asia. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(38), 16046–16050. https://doi.org/10.1073/pnas.0903930106

Robillard, M., & Bahuchet, S. (2012). Les Pygées et les Autres: Terminologie Categorization et Politique. *Journal Des Africanistes*, *82*, 167–192.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Roesch, L. F. W., Lorca, G. L., Casella, G., Giongo, A., Naranjo, A., Pionzio, A. M., … Triplett, E. W. (2009). Culture-independent identification of gut bacteria correlated with the onset of diabetes in a rat model. *The ISME Journal*, *3*(5), 536–548. https://doi.org/10.1038/ismej.2009.5

Romond, M. B., Ais, A., Guillemot, F., Bounouader, R., Cortot, A., & Romond, C. (1998). Cell-free whey from milk fermented with Bifidobacterium breve C50 used to modify the colonic microflora of healthy subjects. *Journal of Dairy Science*, *81*(5), 1229–1235. https://doi.org/10.3168/jds.S0022-0302(98)75683-8

Rook, G. A. W., Martinelli, R., & Brunet, L. R. (2003). Innate immune responses to mycobacteria and the downregulation of atopic responses. *Current Opinion in Allergy and Clinical Immunology*, *3*(5), 337–342. https://doi.org/10.1097/01.all.0000092602.76804.ad

Rosa, B. A., Supali, T., Gankpala, L., Djuardi, Y., Sartono, E., Zhou, Y., … Mitreva, M. (2018). Differential human gut microbiome assemblages during soil-transmitted

helminth infections in Indonesia and Liberia. *Microbiome*, *6*(1). https://doi.org/10.1186/s40168-018-0416-5

Rosenbaum, M., Knight, R., & Leibel, R. L. (2015). The gut microbiota in human energy homeostasis and obesity. *Trends in Endocrinology and Metabolism: TEM*, *26*(9), 493–501. https://doi.org/10.1016/j.tem.2015.07.002

Rosenberg, H. F., Dyer, K. D., & Foster, P. S. (2013). Eosinophils: Changing perspectives in health and disease. *Nature Reviews Immunology*, *13*(1), 9–22. https://doi.org/10.1038/nri3341

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., … Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature*, *555*(7695), 210–215. https://doi.org/10.1038/nature25973

Rowland, I. R., Rumney, C. J., Coutts, J. T., & Lievense, L. C. (1998). Effect of Bifidobacterium longum and inulin on gut bacterial metabolism and carcinogen-induced aberrant crypt foci in rats. *Carcinogenesis*, *19*(2), 281–285.

Russell, J. B., & Rychlik, J. L. (2001). Factors That Alter Rumen Microbial Ecology. *Science*, *292*(5519), 1119–1122. https://doi.org/10.1126/science.1058830

Sairanen, U., Piirainen, L., Gråsten, S., Tompuri, T., Mättö, J., Saarela, M., & Korpela, R. (2007). The effect of probiotic fermented milk and inulin on the functions and microecology of the intestine. *The Journal of Dairy Research*, *74*(3), 367–373. https://doi.org/10.1017/S0022029907002713

Salonen, A., Nikkilä, J., Jalanka-Tuovinen, J., Immonen, O., Rajilić-Stojanović, M., Kekkonen, R. A., … de Vos, W. M. (2010). Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial

and archaeal DNA using mechanical cell lysis. *Journal of Microbiological Methods, 81*(2), 127–134. https://doi.org/10.1016/j.mimet.2010.02.007

Sánchez-Patán, F., Cueva, C., Monagas, M., Walton, G. E., Gibson, G. R. M., Quintanilla-López, J. E., … Bartolomé, B. (2012). In vitro fermentation of a red wine extract by human gut microbiota: Changes in microbial groups and formation of phenolic metabolites. *Journal of Agricultural and Food Chemistry*, *60*(9), 2136–2147. https://doi.org/10.1021/jf2040115

Sands, B. (1998). *Eastern and southn African Khoisan: Evaluating claims of distant linguistic relationships*. Rüdiger Köppe.

Sankararaman, S., Patterson, N., Li, H., Pääbo, S., & Reich, D. (2012). The Date of Interbreeding between Neandertals and Modern Humans. *PLOS Genet*, *8*(10), e1002947. https://doi.org/10.1371/journal.pgen.1002947

Sato, H., Kawamura, K., Hayashi, K., Inai, H., & Yamauchi, T. (2012). Addressing the Wild Yam Question: How Baka Hunter-Gatherers Acted and Lived During Two Controlled Foraging Trips in the Tropical Rainforest of Southeastern Cameroon. *Anthropological Science*, *120*, 129–149. https://doi.org/10.1537/ase.110913

Schandock, F., Riber, C. F., Röcker, A., Müller, J. A., Harms, M., Gajda, P., … Zelikin, A. N. (2017). Macromolecular Antiviral Agents against Zika, Ebola, SARS, and Other Pathogenic Viruses. *Advanced Healthcare Materials*, *6*(23), 1700748. https://doi.org/10.1002/adhm.201700748

Scheinfeldt, L. B., Soi, S., Thompson, S., Ranciaro, A., Woldemeskel, D., Beggs, W., … Tishkoff, S. A. (2012). Genetic adaptation to high altitude in the Ethiopian highlands. *Genome Biology*, *13*(1), R1. https://doi.org/10.1186/gb-2012-13-1-r1

Scheinfeldt, L. B., & Tishkoff, S. A. (2013). Recent human adaptation: Genomic approaches, interpretation and insights. *Nature Reviews Genetics*, *14*(10), 692–702. https://doi.org/10.1038/nrg3553

Scher, J. U., Sczesnak, A., Longman, R. S., Segata, N., Ubeda, C., Bielski, C., … Littman, D. R. (2013). Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *ELife*, *2*. https://doi.org/10.7554/eLife.01202

Schlebusch, C. M., Skoglund, P., Sjödin, P., Gattepaille, L. M., Hernandez, D., Jay, F., … Jakobsson, M. (2012). Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science (New York, N.Y.)*, *338*(6105), 374–379. https://doi.org/10.1126/science.1227721

Schley, P. D., & Field, C. J. (2002). The immune-enhancing effects of dietary fibres and prebiotics. *The British Journal of Nutrition*, *87 Suppl 2*, S221-230. https://doi.org/10.1079/BJNBJN/2002541

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., … Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schnorr, S. L. (2018). Meanings, measurements, and musings on the significance of patterns in human microbiome variation. *Current Opinion in Genetics & Development*, *53*, 43–52. https://doi.org/10.1016/j.gde.2018.06.014

Schnorr, S. L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., … Crittenden, A. N. (2014). Gut microbiome of the Hadza hunter-gatherers. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms4654

Schweinfurth, G. (1873). *Im Herzen Von Afrika*. Leipzig:Brockhaus.

Schwiertz, A., Taras, D., Schäfer, K., Beijer, S., Bos, N. A., Donus, C., & Hardt, P. D. (2010). Microbiota and SCFA in lean and overweight healthy subjects. *Obesity (Silver Spring, Md.)*, *18*(1), 190–195. https://doi.org/10.1038/oby.2009.167

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, *12*(6), R60. https://doi.org/10.1186/gb-2011-12-6-r60

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8). https://doi.org/10.1371/journal.pbio.1002533

Sharma, A., Hill, A., Kurbatova, E., Walt, M. van der, Kvasnovsky, C., Tupasi, T. E., … Cegielski, P. (2017). Estimating the future burden of multidrug-resistant and extensively drug-resistant tuberculosis in India, the Philippines, Russia, and South Africa: A mathematical modelling study. *The Lancet Infectious Diseases*, *17*(7), 707–715. https://doi.org/10.1016/S1473-3099(17)30247-5

Sherrill-Mix, S. (2018). *taxonomizr: Functions to Work with NCBI Accessions and Taxonomy*. Retrieved from https://CRAN.R-project.org/package=taxonomizr

Shimada, Y., Kinoshita, M., Harada, K., Mizutani, M., Masahata, K., Kayama, H., & Takeda, K. (2013). Commensal bacteria-dependent indole production enhances

epithelial barrier function in the colon. *PloS One*, *8*(11), e80604. https://doi.org/10.1371/journal.pone.0080604

Silberbauer, G. (1965). *Report to the Government of Bechuanaland on the Bushmen Survey*. Gaborone, Botswana: Bechuanaland Government Printer.

Silberbauer, G. (1972). The G/wi Bushmen. In M. Bicchieri (Ed.), *Hunters and Gatherers Today* (pp. 271–326). New York: Holt, Rinehart and Winston.

Singh, R. K., Chang, H.-W., Yan, D., Lee, K. M., Ucmak, D., Wong, K., … Liao, W. (2017). Influence of diet on the gut microbiome and implications for human health. *Journal of Translational Medicine*, *15*, 73. https://doi.org/10.1186/s12967-017-1175-y

Skoglund, P., Thompson, J. C., Prendergast, M. E., Mittnik, A., Sirak, K., Hajdinjak, M., … Reich, D. (2017). Reconstructing Prehistoric African Population Structure. *Cell*, *171*(1), 59-71.e21. https://doi.org/10.1016/j.cell.2017.08.049

Smits, S. A., Leach, J., Sonnenburg, E. D., Gonzalez, C. G., Lichtman, J. S., Reid, G., … Sonnenburg, J. L. (2017). Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science*, *357*(6353), 802–806. https://doi.org/10.1126/science.aan4834

Song, S. J., Dominguez-Bello, M. G., & Knight, R. (2013). How delivery mode and feeding can shape the bacterial community in the infant gut. *CMAJ: Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, *185*(5), 373–374. https://doi.org/10.1503/cmaj.130147

Song, S. J., Lauber, C., Costello, E. K., Lozupone, C. A., Humphrey, G., Berg-Lyons, D., … Knight, R. (2013). Cohabiting family members share microbiota with one another and with their dogs. *ELife*, *2*, e00458. https://doi.org/10.7554/eLife.00458

Sonnenberg, G. F., Monticelli, L. A., Alenghat, T., Fung, T. C., Hutnick, N. A., Kunisawa, J., … Artis, D. (2012). Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. *Science (New York, N.Y.)*, *336*(6086), 1321–1325. https://doi.org/10.1126/science.1222551

Sonnenborn, U., & Schulze, J. (2009). The non-pathogenic Escherichia coli strain Nissle 1917 – features of a versatile probiotic. *Microbial Ecology in Health and Disease*, *21*(3–4), 122–158. https://doi.org/10.3109/08910600903444267

Sonnenburg, E. D., Smits, S. A., Tikhonov, M., Higginbottom, S. K., Wingreen, N. S., & Sonnenburg, J. L. (2016). Diet-induced extinctions in the gut microbiota compound over generations. *Nature*, *529*(7585), 212–215. https://doi.org/10.1038/nature16504

Sonnenburg, J. L., & Bäckhed, F. (2016). Diet-microbiota interactions as moderators of human metabolism. *Nature*, *535*(7610), 56–64. https://doi.org/10.1038/nature18846

Spanhaak, S., Havenaar, R., & Schaafsma, G. (1998). The effect of consumption of milk fermented by Lactobacillus casei strain Shirota on the intestinal microflora and immune parameters in humans. *European Journal of Clinical Nutrition*, *52*(12), 899–907.

Speed, D., & Balding, D. J. (2015). Relatedness in the post-genomic era: Is it still useful? *Nature Reviews. Genetics*, *16*(1), 33–44. https://doi.org/10.1038/nrg3821

Spiller, R. (2008). Review article: Probiotics and prebiotics in irritable bowel syndrome. *Alimentary Pharmacology & Therapeutics*, *28*(4), 385–396. https://doi.org/10.1111/j.1365-2036.2008.03750.x

Spyrou, M. A., Bos, K. I., Herbig, A., & Krause, J. (2019). Ancient pathogen genomics as an emerging tool for infectious disease research. *Nature Reviews Genetics*, *20*(6), 323. https://doi.org/10.1038/s41576-019-0119-1

Stagaman, K., Cepon-Robins, T. J., Liebert, M. A., Gildner, T. E., Urlacher, S. S., Madimenos, F. C., … Bohannan, B. J. M. (2018). Market Integration Predicts Human Gut Microbiome Attributes across a Gradient of Economic Development. *MSystems*, *3*(1). https://doi.org/10.1128/mSystems.00122-17

Standaert-Vitse, A., Jouault, T., Vandewalle, P., Mille, C., Seddik, M., Sendid, B., … Poulain, D. (2006). Candida albicans is an immunogen for anti-Saccharomyces cerevisiae antibody markers of Crohn's disease. *Gastroenterology*, *130*(6), 1764–1775. https://doi.org/10.1053/j.gastro.2006.02.009

Stenzel, D. J., & Boreham, P. F. (1996). Blastocystis hominis revisited. *Clinical Microbiology Reviews*, *9*(4), 563–584.

Stewart, J. A., Chadwick, V. S., & Murray, A. (2006). Carriage, quantification, and predominance of methanogens and sulfate-reducing bacteria in faecal samples. *Letters in Applied Microbiology*, *43*(1), 58–63. https://doi.org/10.1111/j.1472-765X.2006.01906.x

Stinson, L. F., Payne, M. S., & Keelan, J. A. (2018). A Critical Review of the Bacterial Baptism Hypothesis and the Impact of Cesarean Delivery on the Infant Microbiome. *Frontiers in Medicine*, *5*. https://doi.org/10.3389/fmed.2018.00135

Stone, A. C., Battistuzzi, F. U., Kubatko, L. S., Perry, G. H., Trudeau, E., Lin, H., & Kumar, S. (2010). More reliable estimates of divergence times in Pan using complete mtDNA sequences and accounting for population structure. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1556), 3277–3288. https://doi.org/10.1098/rstb.2010.0096

Strachan, D. P. (2000). Family size, infection and atopy: The first decade of the "hygiene hypothesis." *Thorax*, *55*(suppl 1), S2-10. https://doi.org/10.1136/thorax.55.suppl_1.S2

Stringer, C., Harvati, K., Allsworth-Jones, P., Grün, R., & Folorunso, C. (2010). New research on the Iwo Eleru cranium from Nigeria. *Am J Phys Anthropol, S50*, 292.

Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C. A., Maza, O., … Elinav, E. (2014). Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature, advance online publication*. https://doi.org/10.1038/nature13793

Świątecka, D., Dominika, Ś., Narbad, A., Arjan, N., Ridgway, K. P., Karyn, R. P., … Henryk, K. (2011). The study on the impact of glycated pea proteins on human intestinal bacteria. *International Journal of Food Microbiology*, *145*(1), 267–272. https://doi.org/10.1016/j.ijfoodmicro.2011.01.002

Tamesse, J. l., Kekeunou, S., Tchouamou, C. L. D., & Meupia, M. j. (2018). Villagers' knowledge of some edible insects in southern Cameroon: Crickets, termites, honeybees and cockchafers. *Journal of Insects as Food and Feed*, *4*(4), 203–209. https://doi.org/10.3920/JIFF2017.0077

Tanaka, J. (1980). *The San, Hunter-Gatherers of the Kalahari*. Tokyo: University of Tokyo Press.

Teklehaymanot, T. (2009). Intestinal parasitosis among Kara and Kwego semipastoralist tribes in lower Omo Valley, Southwestern Ethiopia. *Ethiopian Journal of Health Development*, *23*(1). https://doi.org/10.4314/ejhd.v23i1.44838

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

The World Bank Group. (n.d.). Climate Change Knowledge Portal. Retrieved April 12, 2018, from http://sdwebx.worldbank.org/climateportal/index.cfm?page=country_historical_climate&ThisCCode=TZA

Tilg, H., Adolph, T. E., Gerner, R. R., & Moschen, A. R. (2018). The Intestinal Microbiota in Colorectal Cancer. *Cancer Cell*, *33*(6), 954–964. https://doi.org/10.1016/j.ccell.2018.03.004

Timmann, C., Thye, T., Vens, M., Evans, J., May, J., Ehmen, C., … Horstmann, R. D. (2012). Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature*, *489*(7416), 443–446. https://doi.org/10.1038/nature11334

Timmerman, H. M., Rutten, N. B. M. M., Boekhorst, J., Saulnier, D. M., Kortman, G. A. M., Contractor, N., … Rijkers, G. T. (2017). Intestinal colonisation patterns in breastfed and formula-fed infants during the first 12 weeks of life reveal sequential microbiota signatures. *Scientific Reports*, *7*(1), 8327. https://doi.org/10.1038/s41598-017-08268-4

Tishkoff, S. (2017). Genetics of pigmentation in Eastern and Southern African populations study. *DbGAP, Accession phs001396.v1.p1*. Retrieved from https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001396.v1.p1

Tishkoff, S. A., Gonder, M. K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., … Mountain, J. L. (2007). History of Click-Speaking Populations of Africa Inferred from mtDNA and Y Chromosome Genetic Variation. *Molecular Biology and Evolution*, *24*(10), 2180–2195. https://doi.org/10.1093/molbev/msm155

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., … Williams, S. M. (2009). The genetic structure and history of Africans and African Americans. *Science (New York, N.Y.)*, *324*(5930), 1035–1044. https://doi.org/10.1126/science.1172257

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., … Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, *39*(1), 31–40. https://doi.org/10.1038/ng1946

Tishkoff, S. A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., … Clark, A. G. (2001). Haplotype Diversity and Linkage Disequilibrium at Human G6PD: Recent Origin of Alleles That Confer Malarial Resistance. *Science*, *293*(5529), 455–462. https://doi.org/10.1126/science.1061573

Tito, Raul Y., Knights, D., Metcalf, J., Obregon-Tito, A. J., Cleeland, L., Najar, F., … Lewis, C. M. (2012). Insights from Characterizing Extinct Human Gut

Microbiomes. *PLoS ONE*, *7*(12), e51146. https://doi.org/10.1371/journal.pone.0051146

Tito, Raúl Y., Macmil, S., Wiley, G., Najar, F., Cleeland, L., Qu, C., … Lewis, C. M. (2008). Phylotyping and Functional Analysis of Two Ancient Human Microbiomes. *PLoS ONE*, *3*(11). https://doi.org/10.1371/journal.pone.0003703

Torgerson, P. R., Devleesschauwer, B., Praet, N., Speybroeck, N., Willingham, A. L., Kasuga, F., … de Silva, N. (2015). World Health Organization Estimates of the Global and Regional Disease Burden of 11 Foodborne Parasitic Diseases, 2010: A Data Synthesis. *PLoS Medicine*, *12*(12), e1001920. https://doi.org/10.1371/journal.pmed.1001920

Tormo Carnicer, R., Infante Piña, D., Roselló Mayans, E., & Bartolomé Comas, R. (2006). [Intake of fermented milk containing Lactobacillus casei DN-114 001 and its effect on gut flora]. *Anales De Pediatria (Barcelona, Spain: 2003)*, *65*(5), 448–453.

Torres, S., Clayton, J. B., Danzeisen, J. L., Ward, T., Huang, H., Knights, D., & Johnson, T. J. (2017). Diverse bacterial communities exist on canine skin and are impacted by cohabitation and time. *PeerJ*, *5*. https://doi.org/10.7717/peerj.3075

Towner, J. S., Khristova, M. L., Sealy, T. K., Vincent, M. J., Erickson, B. R., Bawiec, D. A., … Nichol, S. T. (2006). Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola. *Journal of Virology*, *80*(13), 6497–6516. https://doi.org/10.1128/JVI.00069-06

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., … Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, *12*(10), 902–903. https://doi.org/10.1038/nmeth.3589

Turnbaugh, P. J., Bäckhed, F., Fulton, L., & Gordon, J. I. (2008). Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host & Microbe*, *3*(4), 213–223. https://doi.org/10.1016/j.chom.2008.02.015

Turnbaugh, P. J., & Gordon, J. I. (2009). The core gut microbiome, energy balance and obesity. *The Journal of Physiology*, *587*(17), 4153–4158. https://doi.org/10.1113/jphysiol.2009.174136

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., … Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480–484. https://doi.org/10.1038/nature07540

Turnbaugh, P. J., Henrissat, B., & Gordon, J. I. (2010). Viewing the human microbiome through three-dimensional glasses: Integrating structural and functional studies to better define the properties of myriad carbohydrate-active enzymes. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*, *66*(Pt 10), 1261–1264. https://doi.org/10.1107/S1744309110029088

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, *444*(7122), 1027–1031. https://doi.org/10.1038/nature05414

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., & Gordon, J. I. (2009). The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice. *Science Translational Medicine*, *1*(6), 6ra14. https://doi.org/10.1126/scitranslmed.3000322

Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M. S., Kevans, D., Smith, M. I., … Croitoru, K. (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. *Nature Genetics*, *48*(11), 1413–1417. https://doi.org/10.1038/ng.3693

Tzounis, X., Rodriguez-Mateos, A., Vulevic, J., Gibson, G. R., Kwik-Uribe, C., & Spencer, J. P. E. (2011). Prebiotic evaluation of cocoa-derived flavanols in healthy humans by using a randomized, controlled, double-blind, crossover intervention study. *The American Journal of Clinical Nutrition*, *93*(1), 62–72. https://doi.org/10.3945/ajcn.110.000075

Tzounis, X., Vulevic, J., Kuhnle, G. G. C., George, T., Leonczak, J., Gibson, G. R., … Spencer, J. P. E. (2008). Flavanol monomer-induced changes to the human faecal microflora. *The British Journal of Nutrition*, *99*(4), 782–792. https://doi.org/10.1017/S0007114507853384

UNAIDS. (2017). *Ending AIDS: Progress towards the 90-90-90 targets.*

Underhill, D. M., & Iliev, I. D. (2014). The mycobiota: Interactions between commensal fungi and the host immune system. *Nature Reviews Immunology*, *14*(6), 405–416. https://doi.org/10.1038/nri3684

Uren, C., Kim, M., Martin, A. R., Bobo, D., Gignoux, C. R., Helden, D. V., … Henn, B. M. (2016). *Fine-scale human population structure in southern Africa reflects ecological boundaries*. https://doi.org/10.1101/038729

Uronis, J. M., & Threadgill, D. W. (2009). Murine models of colorectal cancer. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, *20*(5), 261–268. https://doi.org/10.1007/s00335-009-9186-5

Urwin, H. J., Miles, E. A., Noakes, P. S., Kremmyda, L.-S., Vlachava, M., Diaper, N. D., … Yaqoob, P. (2014). Effect of salmon consumption during pregnancy on maternal and infant faecal microbiota, secretory IgA and calprotectin. *The British Journal of Nutrition*, *111*(5), 773–784. https://doi.org/10.1017/S0007114513003097

Valiente-Noailles, C. (1993). *The Kua: Life and soul of the Central Kalahari Bushmen*. Rotterdam: A.A. Balkema.

van Heerden, J., Ehlers, M. M., Heim, A., & Grabow, W. O. K. (2005). Prevalence, quantification and typing of adenoviruses detected in river and treated drinking water in South Africa. *Journal of Applied Microbiology*, *99*(2), 234–242. https://doi.org/10.1111/j.1365-2672.2005.02617.x

van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E. G., de Vos, W. M., … Keller, J. J. (2013). Duodenal Infusion of Donor Feces for Recurrent Clostridium difficile. *New England Journal of Medicine*, *368*(5), 407–415. https://doi.org/10.1056/NEJMoa1205037

Vangay, P., Johnson, A. J., Ward, T. L., Al-Ghalith, G. A., Shields-Cutler, R. R., Hillmann, B. M., … Knights, D. (2018). US Immigration Westernizes the Human

Gut Microbiome. *Cell*, *175*(4), 962-972.e10. https://doi.org/10.1016/j.cell.2018.10.029

Vázquez-Castellanos, J. F., Serrano-Villar, S., Latorre, A., Artacho, A., Ferrús, M. L., Madrid, N., … Moya, A. (2015). Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals. *Mucosal Immunology*, *8*(4), 760–772. https://doi.org/10.1038/mi.2014.107

Veech, J. A. (2013). A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, *22*(2), 252–260. https://doi.org/10.1111/j.1466-8238.2012.00789.x

Veeramah, K. R., Wegmann, D., Woerner, A., Mendez, F. L., Watkins, J. C., Destro-Bisol, G., … Hammer, M. F. (2012). An Early Divergence of KhoeSan Ancestors from Those of Other Modern Humans Is Supported by an ABC-Based Analysis of Autosomal Resequencing Data. *Molecular Biology and Evolution*, *29*(2), 617–630. https://doi.org/10.1093/molbev/msr212

Veeramah, Krishna R, & Hammer, M. F. (2014). The impact of whole-genome sequencing on the reconstruction of human population history. *Nature Reviews. Genetics*, *15*(3), 149–162. https://doi.org/10.1038/nrg3625

Vendrame, S., Guglielmetti, S., Riso, P., Arioli, S., Klimis-Zacas, D., & Porrini, M. (2011). Six-week consumption of a wild blueberry powder drink increases bifidobacteria in the human gut. *Journal of Agricultural and Food Chemistry*, *59*(24), 12815–12820. https://doi.org/10.1021/jf2028686

Ventura, M., Canchaya, C., Fitzgerald, G. F., Gupta, R. S., & van Sinderen, D. (2007). Genomics as a means to understand bacterial phylogeny and ecological

adaptation: The case of bifidobacteria. *Antonie Van Leeuwenhoek*, *91*(4), 351–372. https://doi.org/10.1007/s10482-006-9122-6

Verberkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., … Jansson, J. K. (2009). Shotgun metaproteomics of the human distal gut microbiota. *The ISME Journal*, *3*(2), 179–189. https://doi.org/10.1038/ismej.2008.108

Vernot, B., & Akey, J. M. (2014). Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. *Science*, *343*(6174), 1017–1021. https://doi.org/10.1126/science.1245938

Vernot, B., & Akey, J. M. (2015). Complex History of Admixture between Modern Humans and Neandertals. *The American Journal of Human Genetics*, *96*(3), 448–453. https://doi.org/10.1016/j.ajhg.2015.01.006

Vijay-Kumar, M., Aitken, J. D., Carvalho, F. A., Cullender, T. C., Mwangi, S., Srinivasan, S., … Gewirtz, A. T. (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science (New York, N.Y.)*, *328*(5975), 228–231. https://doi.org/10.1126/science.1179721

Vlčková, K., Kreisinger, J., Pafčo, B., Čížková, D., Tagg, N., Hehl, A. B., & Modrý, D. (2018). Diversity of Entamoeba spp. in African great apes and humans: An insight from Illumina MiSeq high-throughput sequencing. *International Journal for Parasitology*, *48*(7), 519–530. https://doi.org/10.1016/j.ijpara.2017.11.008

Vlčková, K., Pafčo, B., Petrželková, K. J., Modrý, D., Todd, A., Yeoman, C. J., … Gomez, A. (2018). Relationships Between Gastrointestinal Parasite Infections and

the Fecal Microbiome in Free-Ranging Western Lowland Gorillas. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.01202

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A Map of Recent Positive Selection in the Human Genome. *PLoS Biology*, *4*(3). https://doi.org/10.1371/journal.pbio.0040072

Vrieze, A., Out, C., Fuentes, S., Jonker, L., Reuling, I., Kootte, R. S., … Nieuwdorp, M. (2014). Impact of oral vancomycin on gut microbiota, bile acid metabolism, and insulin sensitivity. *Journal of Hepatology*, *60*(4), 824–831. https://doi.org/10.1016/j.jhep.2013.11.034

Vrieze, A., Van Nood, E., Holleman, F., Salojärvi, J., Kootte, R. S., Bartelsman, J. F. W. M., … Nieuwdorp, M. (2012). Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology*, *143*(4), 913-916.e7. https://doi.org/10.1053/j.gastro.2012.06.031

Vujkovic-Cvijin, I., Dunham, R. M., Iwai, S., Maher, M. C., Albright, R. G., Broadhurst, M. J., … McCune, J. M. (2013). Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism. *Science Translational Medicine*, *5*(193), 193ra91. https://doi.org/10.1126/scitranslmed.3006438

Wacklin, P., Laurikka, P., Lindfors, K., Collin, P., Salmi, T., Lähdeaho, M.-L., … Kaukinen, K. (2014). Altered duodenal microbiota composition in celiac disease patients suffering from persistent symptoms on a long-term gluten-free diet. *The American Journal of Gastroenterology*, *109*(12), 1933–1941. https://doi.org/10.1038/ajg.2014.355

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., … Flint, H. J. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *The ISME Journal*, *5*(2), 220–230. https://doi.org/10.1038/ismej.2010.118

Wall, J. D., Lohmueller, K. E., & Plagnol, V. (2009). Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Molecular Biology and Evolution*, *26*(8), 1823–1827. https://doi.org/10.1093/molbev/msp096

Wall, J. D., Yang, M. A., Jay, F., Kim, S. K., Durand, E. Y., Stevison, L. S., … Slatkin, M. (2013). Higher levels of neanderthal ancestry in East Asians than in Europeans. *Genetics*, *194*(1), 199–209. https://doi.org/10.1534/genetics.112.148213

Walter, J., & Ley, R. (2011). The Human Gut Microbiome: Ecology and Recent Evolutionary Changes. *Annual Review of Microbiology*, *65*(1), 411–429. https://doi.org/10.1146/annurev-micro-090110-102830

Wammes, L. J., Hamid, F., Wiria, A. E., May, L., Kaisar, M. M. M., Prasetyani-Gieseler, M. A., … Yazdanbakhsh, M. (2016). Community deworming alleviates geohelminth-induced immune hyporesponsiveness. *Proceedings of the National Academy of Sciences*, *113*(44), 12526–12531. https://doi.org/10.1073/pnas.1604570113

Wang, Jingjing, Tang, H., Zhang, C., Zhao, Y., Derrien, M., Rocher, E., … Shen, J. (2015). Modulation of gut microbiota during probiotic-mediated attenuation of

metabolic syndrome in high fat diet-fed mice. *The ISME Journal*, *9*(1), 1–15. https://doi.org/10.1038/ismej.2014.99

Wang, Jun, Thingholm, L. B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J. R., … Franke, A. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nature Genetics*, *48*(11), 1396–1406. https://doi.org/10.1038/ng.3695

Wang, S., Zhu, H., Lu, C., Kang, Z., Luo, Y., Feng, L., & Lu, X. (2012). Fermented milk supplemented with probiotics and prebiotics can effectively alter the intestinal microbiota and immunity of host animals. *Journal of Dairy Science*, *95*(9), 4813–4822. https://doi.org/10.3168/jds.2012-5426

Wang, W., Li, C., Li, F., Wang, X., Zhang, X., Liu, T., … Li, B. (2016). Effects of early feeding on the host rumen transcriptome and bacterial diversity in lambs. *Scientific Reports*, *6*, 32479. https://doi.org/10.1038/srep32479

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., … Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, *46*(4), 336–344. https://doi.org/10.1038/ng.2906

Wei, T., & Simko, V. (2017). *R package "corrplot": Visualization of a Correlation Matrix*. Retrieved from https://github.com/taiyun/corrplot

Wensaas, K.-A., Langeland, N., Hanevik, K., Mørch, K., Eide, G. E., & Rortveit, G. (2012). Irritable bowel syndrome and chronic fatigue 3 years after acute giardiasis: Historic cohort study. *Gut*, *61*(2), 214–219. https://doi.org/10.1136/gutjnl-2011-300220

West, N. P., Christophersen, C. T., Pyne, D. B., Cripps, A. W., Conlon, M. A., Topping, D. L., … Clarke, J. M. (2013). Butyrylated starch increases colonic butyrate concentration but has limited effects on immunity in healthy physically active individuals. *Exercise Immunology Review*, *19*, 102–119.

White, T. D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G. D., Suwa, G., & Howell, F. C. (2003). Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature*, *423*(6941), 742–747. https://doi.org/10.1038/nature01669

Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, *21*(12), 1–20.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Retrieved from http://ggplot2.org

Wickham, H. (2017). *tidyverse: Easily Install and Load the "Tidyverse."* Retrieved from https://CRAN.R-project.org/package=tidyverse

Williams, B., Landay, A., & Presti, R. M. (2016). Microbiome alterations in HIV infection a review. *Cellular Microbiology*, *18*(5), 645–651. https://doi.org/10.1111/cmi.12588

Wilson, I. D., & Nicholson, J. K. (2009). The role of gut microbiota in drug response. *Current Pharmaceutical Design*, *15*(13), 1519–1523.

Wixon, J., & Kell, D. (2000). The Kyoto encyclopedia of genes and genomesâ€ ”KEGG. *Yeast*, *17*, 48–55.

Wolf, A. B., & Akey, J. M. (2018). Outstanding questions in the study of archaic hominin admixture. *PLOS Genetics*, *14*(5), e1007349. https://doi.org/10.1371/journal.pgen.1007349

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3), R46. https://doi.org/10.1186/gb-2014-15-3-r46

Woodmansey, E. J., McMurdo, M. E. T., Macfarlane, G. T., & Macfarlane, S. (2004). Comparison of compositions and metabolic activities of fecal microbiotas in young adults and in antibiotic-treated and non-antibiotic-treated elderly subjects. *Applied and Environmental Microbiology*, *70*(10), 6113–6122. https://doi.org/10.1128/AEM.70.10.6113-6122.2004

World Health Organization. (2006). *Progress on global access to HIV antiretroviral therapy: A report on "3 by 5" and beyond, March 2006*. Retrieved from http://www.who.int/hiv/fullreport_en_highres.pdf

World Health Organization. (2013). *World malaria report: 2013*. Geneva.

World Health Organization, & Global Malaria Programme. (2017). *World malaria report 2017*.

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., … Lewis, J. D. (2011). Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science*, *334*(6052), 105–108. https://doi.org/10.1126/science.1208344

Wu, G. D., Compher, C., Chen, E. Z., Smith, S. A., Shah, R. D., Bittinger, K., … Lewis, J. D. (2014). Comparative metabolomics in vegans and omnivores reveal constraints on diet-dependent gut microbiota metabolite production. *Gut*. https://doi.org/10.1136/gutjnl-2014-308209

Wu, X., Ma, C., Han, L., Nawaz, M., Gao, F., Zhang, X., … Xu, J. (2010). Molecular characterisation of the faecal microbiota in patients with type II diabetes. *Current Microbiology*, *61*(1), 69–78. https://doi.org/10.1007/s00284-010-9582-9

Xie, H., Guo, R., Zhong, H., Feng, Q., Lan, Z., Qin, B., … Jia, H. (2016). Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome. *Cell Systems*, *3*(6), 572-584.e3. https://doi.org/10.1016/j.cels.2016.10.004

Xu, D., Pavlidis, P., Taskent, R. O., Alachiotis, N., Flanagan, C., DeGiorgio, M., … Gokcumen, O. (2017). Archaic Hominin Introgression in Africa Contributes to Functional Salivary MUC7 Genetic Variation. *Molecular Biology and Evolution*, *34*(10), 2704–2715. https://doi.org/10.1093/molbev/msx206

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., … Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569. https://doi.org/10.1038/ng.608

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang, Y.-J., & Sheu, B.-S. (2012). Probiotics-containing yogurts suppress Helicobacter pylori load and modify immune response and intestinal microbiota in the Helicobacter pylori-infected children. *Helicobacter*, *17*(4), 297–304. https://doi.org/10.1111/j.1523-5378.2012.00941.x

Yatsuka, H. (2016). Hunter-Gatherers and their Neighbors in Asia, Africa, and South America. *Senri Ethnological Studies*, *94*, 81–105.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., … Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), 222–227. https://doi.org/10.1038/nature11053

Yooseph, S., Kirkness, E. F., Tran, T. M., Harkins, D. M., Jones, M. B., Torralba, M. G., … Nelson, K. E. (2015). Stool microbiota composition is associated with the prospective risk of Plasmodium falciparum infection. *BMC Genomics*, *16*, 631. https://doi.org/10.1186/s12864-015-1819-3

Yu, Z.-T., Liu, B., Mukherjee, P., & Newburg, D. S. (2013). Trametes versicolor extract modifies human fecal microbiota composition in vitro. *Plant Foods for Human Nutrition (Dordrecht, Netherlands)*, *68*(2), 107–112. https://doi.org/10.1007/s11130-013-0342-4

Yun, Y., Kim, H.-N., Kim, S. E., Heo, S. G., Chang, Y., Ryu, S., … Kim, H.-L. (2017). Comparative analysis of gut microbiota associated with body mass index in a large Korean cohort. *BMC Microbiology*, *17*(1), 151. https://doi.org/10.1186/s12866-017-1052-0

Yurkovetskiy, L., Burrows, M., Khan, A. A., Graham, L., Volchkov, P., Becker, L., … Chervonsky, A. V. (2013). Gender bias in autoimmunity is influenced by microbiota. *Immunity*, *39*(2). https://doi.org/10.1016/j.immuni.2013.08.013

Zaiss, M. M., Rapin, A., Lebon, L., Dubey, L. K., Mosconi, I., Sarter, K., … Harris, N. L. (2015). The Intestinal Microbiota Contributes to the Ability of Helminths to

Modulate Allergic Inflammation. *Immunity*, *43*(5), 998–1010. https://doi.org/10.1016/j.immuni.2015.09.012

Zaura, E., Brandt, B. W., Teixeira de Mattos, M. J., Buijs, M. J., Caspers, M. P. M., Rashid, M.-U., … Crielaard, W. (2015). Same Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome versus Long-Term Microbial Shifts in Feces. *MBio*, *6*(6), e01693-01615. https://doi.org/10.1128/mBio.01693-15

Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., … Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*, *352*(6285), 565–569. https://doi.org/10.1126/science.aad3369

Zhong, Y., Huang, C.-Y., He, T., & Harmsen, H. M. J. (2006). [Effect of probiotics and yogurt on colonic microflora in subjects with lactose intolerance]. *Wei Sheng Yan Jiu = Journal of Hygiene Research*, *35*(5), 587–591.

Zwielehner, J., Liszt, K., Handschur, M., Lassl, C., Lapin, A., & Haslberger, A. G. (2009). Combined PCR-DGGE fingerprinting and quantitative-PCR indicates shifts in fecal population sizes and diversity of Bacteroides, bifidobacteria and Clostridium cluster IV in institutionalized elderly. *Experimental Gerontology*, *44*(6–7), 440–446. https://doi.org/10.1016/j.exger.2009.04.002