Title
Ecology and evolution of the mammalian gut microbiota

Author
Suzuki, Taichi

Publication Date
2018

Peer reviewed|Thesis/dissertation

Ecology and evolution of the mammalian gut microbiota

By

Taichi Suzuki

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael W. Nachman, Chair
Professor Wayne P. Sousa
Professor Mary K. Firestone

Spring 2018

Abstract

Ecology and evolution of the mammalian gut microbiota

by

Taichi Suzuki

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Michael W. Nachman, Chair

Symbiotic microbial communities can affect the health and fitness of the host. Recent studies in humans and mouse models have demonstrated previously unknown roles for the gut microbiota in mammalian digestion, immunity, behavior, and development. These findings suggest that the compositional variation in gut microbiotas may play a fundamental role in host biology. However, the determinants and implications of the gut microbiota in natural populations of mammals are less well understood. My dissertation aims to identify genetic and environmental factors shaping variation in the gut microbiota and to understand the role of gut microbiota in mammalian biology. First, I reviewed studies over the last decade that show links between variation in symbiotic microbiota and variation in host phenotype in natural populations of mammals to motivate the work. Next, I found obesity-associated gut microbial composition tend to be more prevalent in humans living in colder environments consistent with the ecological pattern known as Bergmann's rule. The results suggest a link between the gut microbiota and climatic adaptation. To further investigate whether the pattern observed in humans is general, I conducted a series of studies where I combined field observations and laboratory experiments to investigate the determinants of the gut microbiota and their roles in environmental adaptation using natural populations of house mice *(Mus musculus)* as a model system. I have characterized the gut microbiota of wild house mice using 16S amplicon sequencing within individuals' gastrointestinal tract and between individuals' cecum across two altitudinal and three latitudinal transects across the Americas. Microbiotas of the lower gastrointestinal tract showed greater individual differences compared to the upper gastrointestinal tract. The individual differences in the cecal microbiota were explained by differences in host genetic distance independent of geographic distance. Several gene-bacteria associations were identified from a microbiome genome-wide association study (mGWAS) using exome sequences. The results from the altitudinal and latitudinal transects suggest that differences in partial pressure of oxygen and host body mass may cause changes in gut microbiota. Beneficial functions of the gut microbiota regulating blood pressure at high altitudes and facilitating energy harvest at high latitudes were proposed. Many of the findings here have been replicated in other mammalian systems, including humans. Together, the results suggest general mechanisms governing the assembly and function of the mammalian gut microbiota.

Dedication

To house mice that contributed to science

Table of Contents

Acknowledgements

The work was generously funded by many grants and fellowships, including a Doctorate Dissertation Improvement Grant from the NSF; Young Explorers Grant from National Geographic Society; Jerry O. Wolff Fellowship, Leeper Fund, Karl Koford Fund, Wilhelm L. F. Martens Fund, and Louise Kellogg Fund from the Museum of Vertebrate Zoology (MVZ) at UC Berkeley; Graduate Dean's Summer Research Grant and Graduate Division Summer Grant from the UC Berkeley Graduate Division; Galileo Circle Scholarship from College of Science at the University of Arizona; and Grant-in-Aid of Research from Sigma Xi. The work was also supported by NSF and NIH grants to Michael Nachman.

I would first like to thank my advisor and committee chair, Dr. Michael Nachman, for being the best mentor, teacher, and researcher supporting my research and personal development as a scientist throughout the years. His paper on rock pocket mice made me decide to apply to his lab and his professional attitude and dedication to science made me decide to pursue a career in academia. I would also like to thank my dissertation committees, Dr. Wayne Sousa and Dr. Mary Firestone. Wayne taught me community ecology and parasitology by two-hour long weekly meetings for a semester and gave me personal support. Mary taught me microbiology, and her patience, encouragement, and intellectual guidance supported my research. I would also like to thank my collaborators and colleagues that played a significant role in my dissertation research and graduate student life, including Dr. M. Iwasa who plays a role as a mentor in my life, Dr. M. Worobey introduced me to microbial ecology and a rotation project with him sparked the current dissertation idea, Dr. V. Rich, T. O'Conner, Dr. E. Vo, Dr. D. Armintage, and Dr. K. Bi provided assistance in computational analyses and useful discussions, Dr, M. Edwards, I. Moya, and all other international collaborators helped collect the mouse samples necessary for the work, Dr. R. Bowie, Dr. R. Nielsen, Dr. M. Shapira, Dr. H. Ochman, Dr. N. Moran, Dr. N. Whiteman, Dr. J. Bronstein, Dr. T. Taylor, Dr. Y. Zhang, Dr. H-T. Yu, Dr. J. Gordon, Dr. R. Ley, Dr. W. Hood, Dr. D. Dearing, Dr. K. Kohl, Dr. A. Chandler, Dr. J. Sanders, Dr. H. Carey, Dr. K. Duddleston, Dr. C. Moreau, Dr. S. Bordenstein, Dr. C. Phillips, Dr. S. Cohen, Dr. M. Noor, and members of the Gut Club and Eco Lunch, provided valuable comments on the work.

I also would like to express my gratitude to the members of the Nachman lab and the MVZ community for treating me as part of an academic family and providing a supportive environment. Past and current members of the Nachman lab played a critical role on the development of the dissertation and personal growth, including Dr. B. Payseur, Dr. J. Good, Dr. M. Dean, Dr. G. Wlasiuk, Dr. T. Salcedo, M. Beckwith, P. Liebig, J. Jonas, Dr. P. Campbell, Dr. M. Phifer-Rixey, M. Bomhoff, N. Bittner, T-T. Lin, K. Mack, Dr. M. Sheehan, Dr. F. Martins, Dr. A. Chavez, Dr. C. Emerling, Dr. K. Ferris, Dr. G. Bradburd, Dr. A. Moeller, M. Ballinger, and S. Banker. I also thank the undergraduate assistants in the Nachman lab for all their effort, including G. Heyer, J. Yang, Y. Sato, W. Feng, M. Dzima, A. Petrosky, S. Berkley, J. Zhang, E. Shiah, J. Godoy, and M. Choudhery. I also thank the MVZ community, including L. Smith who supported the wet-lab work, Dr. C. Conroy and T. Wong who helped me with field

supplies and sample shipments, Dr. J. Patton and Dr. D. Wake who have been always open to chat about science, and all the MVZ graduate students, postdocs, staff, and curators that provided a supportive environment to pursue the graduate program as a MVZ student.

Finally, I thank my family and friends that gave me support to pursue my career and dream. Special thanks to Dozer and Miso (dogs) for being my study partners to complete this dissertation.

Chapter 1

Links between Natural variation in the microbiome and host fitness in wild mammals

This chapter has been previously published and is reproduced here in accordance with the journal's article sharing policy:

Suzuki, T.A. 2017. Links between natural variation in the microbiome and host fitness in wild mammals. Integrative and Comparative Biology.
DOI: 10.1093/icb/icx104

Abstract

Recent studies in model organisms have shown that compositional variation in the microbiome can affect a variety of host phenotypes including those related to digestion, development, immunity, and behavior. Natural variation in the microbiome within and between natural populations and species may also affect host phenotypes and thus fitness in the wild. Here, I review recent evidence that compositional variation in the microbiome may affect host phenotypes and fitness in wild mammals. Studies over the last decade indicate that natural variation in the mammalian microbiome may be important in the assistance of energy uptake from different diet types, detoxification of plant secondary compounds, protection from pathogens, chemical communication, and behavior. I discuss the importance of combining both field observations and manipulative experiments in a single system to fully characterize the functions and fitness effects of the microbiome. Finally, I discuss the evolutionary consequences of mammal-microbiome associations by proposing a framework to test how natural selection on hosts is mediated by the microbiome.

1.1. Introduction

Recent studies in animal models have shown that variation in the microbiome can affect the host phenotype in many ways, including through traits related to digestion, development, immunity, and behavior (McFall-Ngai *et al.* 2013). These findings suggest that compositional differences in symbiotic microbial communities may play a fundamental role in host ecology and evolution. Understanding the beneficial effects of the mammalian microbiome in natural populations is particularly important for understanding human health as well as for generalizing the findings from lab-reared models and domestic animals. Although the importance of microbes in host health has been well-established mainly using culturing methods, we know less about how natural variation in the microbiome can affect host fitness in wild mammals.

The idea that the microbes may play a beneficial role in host fitness is not new. For example, one of the best studied beneficial functions of the mammalian microbiome is its role in the digestion of plant materials, such as cellulose. In the late 1800s,

microbial fermentation was proposed as a mechanism of cellulose digestion (Zuntz 1879) and was supported empirically (Tappeiner 1884). In the 1940s, protozoa (Hungate 1942, 1943) and bacteria (Hungate 1947; Sijpesteijn 1948) with capabilities to digest cellulose were isolated from the stomach of ruminants. Fermentation products, such as volatile fatty acids, have been shown to be absorbed by the host from the rumen wall into the blood system as energy sources (Barcroft, Mcanally and Phillipson 1944). The concentration of these fatty acids is highest in the fermentation chambers compared to the rest of the gastrointestinal tract in herbivores (Elsden *et al.* 1946). These findings led to classic reviews and books on gut microbiota and digestion (Hungate 1966; Van Soest 1994; Stevens and Hume 1995; Hume 1999).

In the last decade or so, the advancement of DNA sequencing technology has made it possible to investigate variation of the entire microbial community (which was impossible to study based on culturing methods alone). For example, the host-associated microbiome varies between species as well as within and between natural populations. Interspecific variation in microbiomes exists both in captivity (Ley *et al.* 2008a; Muegge *et al.* 2011) and in the wild (Ochman *et al.* 2010; Phillips *et al.* 2012). Between-population variation in the microbial composition has been observed in rodents (Linnenbrink et al. 2013; Kohl et al. 2014b; Kreisinger et al. 2015) and primates (Amato et al. 2013; Moeller et al. 2013a) including humans (Suzuki and Worobey 2014). Within-population variation in the microbiome has also been observed in humans (Arumugam *et al.* 2011), chimpanzees (Moeller *et al.* 2012), and house mice (Wang *et al.* 2014). Most of the work has focused on the gut microbiome, but studies have also described microbes associated with other body sites of the host (Costello *et al.* 2009; Huttenhower *et al.* 2012).

In principle, variation in the microbiome could have a positive effect, a negative effect, or no effect on host fitness. However, it appears unlikely that compositional differences in microbiome would have no impact on host fitness given the known role of microbes in digestion (Hungate 1966; Van Soest 1994; Stevens and Hume 1995; Hume 1999), the abundant biomass of symbiotic microbes associated with hosts (Sender et al. 2016), the known association between microbes and host immunity (Round and Mazmanian 2009a), and the observations that even rare microbes are suggested to play a large role in microbial ecosystem functions (reviewed in Jousset et al. 2017). Nevertheless, it is possible that some rare and transient microbes have little or no effect on host fitness. The current examples of positive and negative effects of microbes on host fitness are sometimes restricted to studies of a single bacterial lineage. For example, the positive effects of bacteria in sap-sucking insects are well-established including the role of bacteria in providing essential nutrients (Baumann 2005), defense from pathogens (Oliver *et al.* 2003), and tolerance to heat (Dunbar *et al.* 2007). The negative effects of microbes on host fitness have also been studied, primarily in the context of microbial pathogens including protists (Allison 1964), bacteria (Lowy 1998), and viruses (Piertney and Oliver 2006). However, fitness effects of variation in an entire microbial community are not well understood. Understanding the association between variation of the microbiome and variation of host fitness requires molecular tools to characterize differences in microbial community composition in natural populations.

Here, I review studies over the last decade investigating inter- and intraspecific variation in the mammalian microbiome that suggest a link between variation in the

microbiome and variation in host phenotype and thus fitness (Table 1). Although most of the evidence is correlational, these studies suggest that natural variation in host-associated microbiomes play a role in mammalian host biology including aspects of digestion, detoxification, immunity, and behavior. I then identify potential future research directions, including the need for manipulative experiments.

1.2. Literature review

1.2.1. Assistance of energy uptake from diet

A series of studies by Gordon and colleagues using fecal transplant experiments in mouse models showed that the gut microbiome can affect fat storage and energy extraction from the diet (Bäckhed *et al.* 2004; Turnbaugh *et al.* 2006). Distantly related mammalian host species kept in zoos show convergence in the microbiome by diet types despite the significant effects of host phylogeny and numerous other factors affecting the composition of the microbial community (Ley *et al.* 2008a; Muegge *et al.* 2011). Furthermore, differences in the microbiome may reflect functional differences in digestion. For example, the microbiomes of mammalian herbivores are enriched for microbial genes that synthesize amino acids compared to those of carnivores which are enriched for microbial genes that degrade amino acids (Muegge *et al.* 2011). Together, the convergence in microbiomes in distant mammalian groups that share a similar diet suggests that the microbiome may play a fundamental role in the ability of mammals to specialize on a particular source of food.

Similar patterns of convergence of microbiomes associated with dietary types have been observed in wild mammals. For example, convergence in microbiomes has been observed among myrmecophagous mammals such as anteaters, aardvarks, and aardwolves that feed on ants and termites (Delsuc *et al.* 2014), among bat species that share similar diet types (Phillips *et al.* 2012; Carrillo-Araujo *et al.* 2015), between chimpanzees and gorillas that share a large fraction of fruit diet (Moeller et al. 2013a), and between carnivorous marine mammals (Nelson et al. 2013a; Nelson 2013b; Soverini et al. 2016). Population-level differences in the microbial composition of humans were also associated with differences in the amount of dietary fiber (De Filippo *et al.* 2010; Ou *et al.* 2013; Schnorr *et al.* 2014; Angelakis *et al.* 2016). Although these studies are consistent with the idea that convergent microbiomes are involved in diet-specific digestion, there are alternative explanations. For example, some of the observed patterns might be explained by common microbes in the shared diet or co-variation with other aspects of the shared environment (Moeller et al. 2013a; Suzuki and Worobey 2014). Thus, convergence of microbiomes among distant mammalian lineages does not, by itself, provide definitive evidence of the beneficial effects of the microbiome on the host. In fact, microbes can compete for resources with the host when the nutrients, such as simple sugars and iron, can be utilized by both hosts and microbes (Wasielewski et al. 2016).

The best evidence of the beneficial role of microbes in energy extraction comes from components of the diet that cannot be utilized directly or easily by the host, such as cellulose. Mammalian herbivores have evolved two major mechanisms for breaking

down plant polysaccharides; foregut fermentation (as in ruminants) and hindgut fermentation (Stevens 1995; Feldhamer et al. 2007). In the case of foregut fermenters, the stomach and/or esophagus are modified and enlarged, while in hindgut fermenters, the main site of fermentation can be either the large intestine or the enlarged cecum. Foregut and hindgut fermenters each have distinct microbial communities that contain high densities of microbes which aid in digestion (Ley *et al.* 2008a; Muegge *et al.* 2011). As mentioned briefly above, the role of microbes in ruminant digestion has been studied extensively in large part due to its economic and agricultural importance (Hungate 1966; Van Soest 1994; Stevens and Hume 1995; Hume 1999). This is now being followed up using molecular techniques (Mccann et al. 2014). Culture-based studies have suggested a role for protozoa, bacteria, and fungi in plant digestion not only in domestic animals, but also in various wild herbivores (Sahu and Kamra 2002). Shotgun metagenomic studies have revealed that microbiomes of herbivorous mammals contain bacterial genes involved in fiber fermentation (Pope et al. 2010; Zhu et al. 2011; Fang et al. 2012).

Species-rich microbial communities may be important for digestion of plant fibers. From interspecific comparisons, the gut microbial communities of herbivores often have the highest species richness (i.e. alpha-diversity) compared to those of other dietary types (Ley *et al.* 2008a; Muegge *et al.* 2011; Phillips *et al.* 2012). An exception to this pattern is the giant panda which has one of the lowest alpha-diversity measures across mammals (Ley et al. 2008; Xue et al. 2015). Although microbial genes involved in cellulose degradation and lignin oxidation were identified in panda feces (Zhu *et al.* 2011; Fang *et al.* 2012), their poor digestion rate may be due to deficiencies in cellulose-digesting bacteria and low alpha-diversity (Xue et al. 2015). Interestingly, even within an individual's gastrointestinal tract, there appears to be an association between alpha-diversity and plant fermentation. The foregut and hindgut fermentation chambers have greater alpha-diversity compared to the small intestine in multiple species of wild-caught rodents (Kohl et al. 2014a; Lu et al. 2014; Suzuki and Nachman 2016). The stability-diversity relationship (Tilman and Downing 1994), an observation that species-rich communities are more stable, resilient, and recover faster from disturbance, has been suggested as an explanation for the species-rich communities observed in herbivores' fermentation chambers (Lu *et al.* 2014; Suzuki and Nachman 2016).

Chitin is the second most abundant biopolymer on the planet and common in the cell walls of fungi and the exoskeletons of arthropods. Although chitin can be broken down by both mammalian and microbial enzymes, microbes may play a particularly important role in chitin degradation (Herwig *et al.* 1984; Simůnek *et al.* 2001; Delsuc *et al.* 2014; Sanders *et al.* 2015). Chitinase-producing bacteria have been isolated in insectivorous bats (Whitaker et al. 2004) as well as in herbivorous and omnivorous mammals that are not adapted uniquely to a chitin-rich diet (Simůnek *et al.* 2001). Chitinase activity has also been detected in the stomach of nine-banded armadillos (Smith et al. 1998) and the intestines of insectivorous bats (Whitaker, Dannelly and Prentice 2004). The involvement of microbes in chitin degradation has been suggested in myrmecophagous mammals that show convergence in their microbial composition (Delsuc *et al.* 2014). Baleen whales also feed on prey that is rich in chitin, and their foregut stomach has been suggested to play a role in the microbial degradation of chitin (Herwig *et al.* 1984). In fact, a recent metagenomic study found an enrichment of bacterial genes associated with chitin degradation in baleen whales (Sanders *et al.* 2015).

Seasonal changes in diet and gut microbial composition in wild mammals also support the involvement of the microbiome in energy extraction. In multiple species of primates (Amato *et al.* 2014; Fogel 2015; Gomez *et al.* 2015; Sun *et al.* 2016), wood mice (Maurice *et al.* 2015), and reindeer (Salgado-Flores *et al.* 2016), microbial composition was found to vary seasonally. The seasonal change in the microbial composition was correlated with the changes in the metabolic profiles in gorilla (Gomez *et al.* 2015) and howler monkeys (Amato *et al.* 2014). Hibernation and torpor are perhaps the most extreme cases of seasonal shifts in diet, requiring both hyperphagia and fasting (Carey et al. 2003). The microbiota is known to differ between active seasons and hibernation periods in thirteen-lined ground squirrels (Carey et al. 2013; Dill-Mcfarland et al. 2014) and arctic ground squirrels (Stevenson et al. 2014). Microbially-provided energy sources such as short-chain fatty acids (SCFAs) have been suggested to serve as a source of fuel during hibernation by bacteria degrading host-derived polysaccharides such as mucins (Carey et al. 2013). Although hibernation reduces the total amount of cecal SCFAs, the relative proportion of acetate (a specific SCFA) increases during hibernation compared to active seasons in the two species of ground squirrels (Carey et al. 2013; Stevenson et al. 2014). The increase in acetate-producing bacteria such as *Akkermansia mucinphila* that degrades mucins (Derrien *et al.* 2008) was observed in fasting and hibernating ground squirrels (Carey et al. 2013; Stevenson et al. 2014) and fasting Syrian hamsters (Sonoyama *et al.* 2009).

Together, these studies are consistent with the hypothesis that variation in the gut microbiome is associated with the variation in energy extraction from diverse diets in mammalian hosts.

1.2.2. Detoxification of plant dietary toxins

Plants defend themselves from herbivores by producing plant secondary compounds or dietary toxins. Thus, detoxifying dietary toxins is a critical challenge for herbivores specializing on plant diets. Mammalian hosts often rely on microbes to detoxify plant toxins. For example, culture-based studies have isolated dietary toxin-degrading bacteria from a wide range of mammalian hosts including various domestic species (Osawa and Sly 1992; Nemoto *et al.* 1995), marsupials (Osawa 1990; Osawa and Sly 1992; Nemoto *et al.* 1995), Ethiopian ruminants (Ephraim et al. 2005), and rodents (Sasaki et al. 2005; Dai et al. 2014; Miller et al. 2014; Kohl et al. 2016b). The occurrence of toxin degrading bacteria was associated with mammals that consume tannin rich diet (Osawa and Sly 1992), but not in others (Nemoto *et al.* 1995). Although detoxification enzymes can be produced both by the host tissue and the microbes, microbial enzyme activity per unit protein greatly exceeds that of the host in spiny mice (Kohl *et al.* 2016a) consistent with the findings in lab rats (Nakano and Gregory 1995). Two species of woodrats that consume dietary toxins in the wild converged on microbial community composition when they were fed dietary toxins in a common environment (Kohl et al. 2012). Dietary toxin feeding experiments in Japanese large wood mice demonstrated that individuals that have prior experience to dietary toxins were associated with a greater abundance of toxin-degrading bacteria and have better performance measured by weight change over time (Shimada *et al.* 2006).

A series of elegant papers by Kohl and Dearing used manipulative experiments to show that microbes aid woodrats in detoxifying plant secondary compounds (reviewed in Kohl and Dearing 2016). Microbes cultured from woodrat guts can degrade plant secondary compounds (Miller et al. 2014; Kohl et al. 2016b), and the ability of woodrats to consume dietary toxins is impaired when rats are given antibiotics (Kohl et al. 2014b). When feces of woodrat donors that regularly consumed dietary toxins were fed to woodrat recipients that were naive to the dietary toxin, the recipients showed an increased ability to feed on the dietary toxin (Kohl et al. 2014b). Similarly, microbial transplant experiments from woodrats to laboratory rats increased the ability of laboratory rats to consume tannic acids (Kohl et al. 2016b). Together, these studies provide strong evidence that differences in the gut microbiome in natural populations are involved in detoxification of dietary toxins of herbivores.

1.2.3. Pathogen defense and development of immune system

While the microbiome clearly plays an important role in digestion and detoxification, another well-studied role is its effect on host immunity (Round and Mazmanian 2009a). In model organisms, the microbiota has been shown to benefit hosts by occupying the niche space of pathogens or by priming the development of the immune system. Recently, viruses in the gut mucus layer were suggested to play a role in protection against bacterial pathogens (Barr *et al.* 2013). Understanding the role of the microbiome in host immunity in wild mammals is another growing research area.

The microbiome of a host may help provide protection from lethal pathogens in natural populations. For example, white-nose syndrome is an emerging infectious disease in North American hibernating bats caused by the fungus, *Pseudogymnoascus destructans* (Warnecke *et al.* 2012). Recent studies isolated bacteria from the skin of healthy bats and showed anti-fungal effects against *P. destructans,* suggesting a beneficial role of skin microbes in pathogen defense (Hoyt et al. 2015; Hamm et al. 2017). A similar link has been identified in chytrid fungus infections in amphibians (reviewed in Rollins-Smith et al. 2011; Colombo et al. 2015). Thus, natural variation in the non-pathogenic microbial community may affect the susceptibility and transmission of fungal diseases in wild mammals.

Infections of viruses and macroparasites are also associated with differences in the gut microbiota in wild mammals. Immunodeficiency virus infection (e.g. HIV in case of humans) was associated with changes in the microbiota of humans (reviewed in Salas and Chang 2014; Williams et al. 2016), chimpanzees (Moeller et al. 2013b), and domestic cats (Weese *et al.* 2015), but not in gorillas (Moeller *et al.* 2015). After HIV infection in humans, gut microbiomes may affect the progression of the disease (Vujkovic-Cvijin *et al.* 2013). Microbial communities seem to respond differently towards different infectious agents. For example, in wild-caught house mice, the degree of viral infection was positively correlated with alpha-diversity of the gut microbial community, whereas the degree of nematode and mite infection was negatively correlated with the alpha-diversity (Weldon *et al.* 2015). Aberrant immune gene expression and intestinal histopathology in hybrid house mice were associated with changes in the microbial community composition compared to the two parental species (Wang *et al.* 2015). Associations between helminth infection and the microbial composition were

observed in yellow-necked mice, although the effect size was very low (Kreisinger *et al.* 2015). These correlations between host immunity and the microbial composition in wild mammals are interesting, but the correlations alone cannot disentangle the cause and the consequence of the association. Nor is it always clear whether these associations reflect positive or negative effects on host fitness. Manipulative experiments including infection trials are necessary to understand the role of microbial communities in host immunity.

1.2.4. Modification of chemical communication and host behavior

Emerging evidence suggests that the microbiome can alter the behavior of mammalian hosts in two ways; (1) changes in olfactory signaling by microbial communities in the scent glands (Ezenwa and Williams 2014), and (2) changes to the central nervous system by microbial communities in the gut (Sampson and Mazmanian 2015). The relationship between the host behavior and the microbiome can be reciprocal in the sense that social interactions between hosts can affect variation in the microbiome (Archie and Tung 2015; Tung et al. 2015; Moeller et al. 2016a). Using computer models, Lewin-Epstein et al (2017) suggested that microbes might play a role in the evolution of host altruism. The implications of the microbiome in behavioral ecology have been reviewed elsewhere (Archie and Theis 2011; Archie and Tung 2015).

In mammals, olfactory communication plays an important role in various behaviors, including mate preference and individual and species recognition (Ezenwa and Williams 2014). The fermentation hypothesis of chemical recognition is the idea that microbial communities in mammalian scent glands play an important role in the production of volatile odors (Albone *et al.* 1974; Albone and Perry 1976; Gorman 1976; Albone 1984). There is now evidence for microbe-produced signals in a variety of animals (reviewed in Ezenwa & Williams, 2014). In wild mammals, the microbiomes in scent glands are known to differ among closely related species (Theis *et al.* 2013), social groups (Theis et al. 2012; Theis et al. 2013; Leclaire et al. 2014), individuals (Gorman 1976; Merritt *et al.* 1982) and individuals with different sex (Gassett et al. 2000; Alexy et al. 2003; Voigt et al. 2005; Theis et al. 2013; Leclaire et al. 2014), reproductive status (Sin *et al.* 2012; Li *et al.* 2016a), and age (Sin et al. 2012; Leclaire et al. 2014). Furthermore, bacteria-mediated odor significantly altered the outcomes of choice tests in Indian mongooses (Gorman 1976) and Brazilian guinea pigs (Zechman *et al.* 1984), consistent with findings in lab mice (Li *et al.* 2012).

Involvement of microbes in olfactory signaling has been observed in humans as well. Humans exhibit MHC-dependent mate preference and kin recognition based on odors (Wedekind *et al.* 1995; Wedekind and Füri 1997; Weisfeld *et al.* 2003). Axillary odors are known to be associated with microbial composition (Leyden *et al.* 1981) and there is evidence that microbes transform odorless host axillary secretions into volatile odors (James *et al.* 2013). Bacteria-produced odorant was more similar between monozygotic twins compared to unrelated individuals (Kuhn and Natsch 2009) supporting the potential involvement of axillary microbes in inbreeding avoidance and kin recognition in humans. The behavioral, hormonal, and neurochemical effects of gut microbiomes through the gut-brain axis has been studied in humans and lab mice including the role of the microbiota on anxiety-like behavior, depression-like behavior, Parkinson's disease, Alzheimer's disease, and autism spectrum disorder (reviewed in

Sampson and Mazmanian 2015). However, outside of humans and lab mice, the implications of the gut-brain axis in the ecology and evolution of wild mammals are largely unexplored.

1.2.5. Manipulative experiments are needed

Overall, the studies presented above are consistent with the idea that natural variation in the microbiome can affect differences in host fitness. One of the advantages of omics-approaches (metagenomics, transcriptomics, proteomics, and metabolomics) is the ability to connect the compositional variation in the microbiome to functional variation in the microbiome, and eventually to the differences in host phenotypes and fitness. However, most of the current evidence is correlative (Table 1). Manipulative experiments in controlled settings are needed to identify causal relationships.

The effects of the microbiome on host phenotype have commonly been tested in model organisms mainly using two approaches; (1) measuring the phenotype of the host without the microbiome by knocking-down the microbial community with antibiotics or by using germ-free animals, or (2) measuring the phenotype of the host with a modified microbiome through transplant experiments, either using bacterial isolates or a whole bacterial community. Although the use of antibiotics has been criticized by having many confounding effects (reviewed in Lundberg et al. 2016), there are several advantages over germ-free models such as studying the depletion of microbiome at different developmental stages and the ability to target bacterial groups by using different classes of antibiotics. While germ-free models also have their own limitations, germ-free models still seem to be the best system for microbial transplant experiments (Lundberg et al. 2016). For example, germ-free lab mice exhibit a variety of abnormalities including the development of the gut, immune system, and brain (Martín et al. 2016). Fecal transplant experiments have demonstrated a causal role of the microbiome in fat storage (Turnbaugh *et al.* 2006), anxiety-like behavior (Heijtz *et al.* 2011) and priming the immune system (Olszak *et al.* 2012). Conducting manipulative experiments to test hypotheses generated by field observations would greatly increase our understanding of the functions of the compositional variation of the microbiome in wild mammals (Kohl et al. 2014b; Brooks et al. 2016; Kohl et al. 2016b).

A particularly powerful approach would be to combine field observations and manipulative experiments in a single species. In human microbiome research, laboratory mice are often used as a stand-in for human subjects because conducting manipulative experiments in humans can be difficult or impossible (Nguyen *et al.* 2015). However, results from mice may not translate easily to humans due to differences in anatomy, physiology, and genetics (Nguyen *et al.* 2015) as well due to the existence of species-specific bacterial lineages (Moeller et al. 2016b) and communities (Ochman *et al.* 2010; Phillips *et al.* 2012; Brooks *et al.* 2016). A system that can combine both observations in natural populations and manipulations in the lab would be useful for assessing the role of microbiome variations on host fitness. Rodents are a particularly tractable group for combining these two approaches for microbiome research in general (Kohl et al. 2014b; Brooks et al. 2016; Kohl et al. 2016b).

1.2.6. Evolutionary consequence of microbiome functions

Evolutionary change in a host that is driven by natural selection could be mediated by microbiome in cases where microbes are faithfully transmitted from one generation to the next. The implications of the fitness effects of the microbiome for host evolution have been widely discussed in the literature (Margulis and Fester 1991; Rosenberg et al. 2007; Zilber-Rosenberg and Rosenberg 2008; McKnite et al. 2012; Bordenstein and Theis 2015; Moran and Sloan 2015). Here, I propose a framework for testing for evidence of host evolution mediated by the microbiome.

The three components of natural selection can be applied to the microbiome; variation, differential success, and inheritance (Fig. 1). First, in order for natural selection to act on the microbiome, variation in the microbiome is required. Second, some of the variation in the microbiome must affect the host fitness. Lastly, inheritance of the microbiome associated with the host fitness is also required, where the offspring microbiome resembles the parent microbiome. Providing evidence of each of the three components in a single system will support the idea that natural selection on hosts is mediated by the microbiome. However, resemblance of the microbiome among related individuals can be generated by a variety of processes and the mechanism is largely unexplored especially in natural populations.

Resemblance of the microbiome between parent and offspring can occur through genetic and non-genetic mechanisms. For example, reciprocal transplant experiments between zebrafish and mice have demonstrated that the host genotype can assemble species-specific microbial communities (Rawls *et al.* 2006). Genes with immune and behavioral functions are known to affect the microbiome from gene knockout mouse strains (Spor et al. 2011). In contrast, the mammalian offspring can acquire beneficial microbes either vertically (e.g. mother's vaginal canal, milk, skin) or horizontally (e.g. environment) without host genotype effects. For example, rodents and lagomorphs often acquire microbes through coprophagy (e.g. eat mother's feces), whereas young ungulates commonly consume soil to acquire microbes (Feldhamer *et al.* 2007). Some insect hosts can acquire beneficial bacteria that confer insecticide resistance from the soil every generation (Kikuchi *et al.* 2012). In lab mice, exposure to a different temperature resulted in different assemblages of microbiome that improves host energy metabolism by plasticity, without any genotypic differences (Chevalier *et al.* 2015; Zięta*k et al.* 2016). Therefore, studying the degree and the stability of the inheritance of the beneficial microbiome including genetic and non-genetic mechanisms is important to understand the evolutionary outcomes of the host.

Finally, a major goal in evolutionary genetics is to link genotypes to phenotypes that affect fitness. Identifying the genetic basis of host phenotypes (e.g. immunity, behavior) that interact with the microbiome is important for understanding how the host genome might regulate the functions of the microbiome (Spor et al. 2011). Host genomic regions that associate with the microbial composition have been identified in genome-wide association studies in humans (reviewed in Goodrich et al. 2016b) and quantitative trait locus mapping studies in mice (Benson *et al.* 2010a; McKnite *et al.* 2012; Srinivas *et al.* 2013; Leamy *et al.* 2014b; Wang *et al.* 2015). Conversely, "heritable" microbial taxa have been identified and can have significant effects on host phenotype such as weight gain (Goodrich et al. 2014, 2016a). Linking host genotype to microbiome variation that affects host fitness will help to understand the evolution of host-microbial interactions.

1.3. Chapter 1 Table

Table 1. Suggestive evidence of compositional and functional variation of the microbiome affecting the host phenotype in natural or semi-natural populations of mammals using culture-independent techniques. See text for culture-based studies and functional studies for further evidence on microbiome functions.

Host phenotype		Mammalian host species		Links between variations of the microbiome and the host phenotype	References
Energy uptake from diet	Plant fiber digestion	Humans	*Homo sapiens*	Population differences in fiber-rich diet were associated with differences in the microbiota in a way that is consistent with the findings in experimental studies.	De Filippo et al. 2010, Ou et al. 2013, Schnorr et al. 2014, Angelakis et al. 2016
		Tammar Wallaby*	*Macropus eugenii*	The foregut microbiome varied compositionally from other herbivores, but the microbiome converged functionally in plant fiber breakdown pathways using metagenomics.	Pope et al. 2010
		Giant panda*	*Ailuropoda melanoleuca*	Genes involved in cellulose metabolism and lignin oxidation were identified, but the microbiome differs compositionally from other herbivores and similar to carnivorous bears.	Zhu et al. 2011, Fang et al. 2012, Xue et al. 2015
		Flying squirrel	*Petaurista alborufus lena*	Convergence in microbiota of fermentaion chambers along the gastrointestinal tract was observed (e.g. fermentation chambers tend to have greater alpha-diversity compared to the small intestines regardless of the presence or absence of the foregut chambers).	Lu et al. 2014
		White-throated woodrat	*Neotoma albigula*		Kohl et al. 2014a
		House mouse	*Mus musculus*		Suzuki et al. 2016
	Chitin digestion	Baleen Whales	Mysticeti sp.	The microbial composition is distinct from other mammals and enriched in bacterial genes associated with chitin degradation using metagenomics.	Sanders et al. 2015
		Myrmecophagous mammals*	(e.g. echidna, aardvark, anteater, tamandua, armadillo, ardwolf, etc.)	Convergence of microbiota was observed across a wide range of myrmecophagous mammals.	Delsuc et al. 2014
	Fish digestion	Marine mammals*	*Mirounga leonina, Hydrurga leptonyx, Dugong dugong*	Convergence of microbiota was observed among marine carnivorous, distinct from the rest of marine and terrestrial mammals.	Nelson et al. 2013a, 2013b
		Bottlenose dolphins*	*Tursiops truncatus*	Convergence of microbiota was observed between bottlenose dolphins and carnivores fish that share precocious diet.	Soverini et al. 2016
	Fruit digestion	Chimpanzee and Gorilla	*Pan and Gorilla*	Convergence of microbiota was observed between chimpanzees and gorillas that share large fraction of a fruit diet.	Moeller et al. 2013a

* Captive individuals in semi-natural environment were used mostly or entirely.

Table 1. Continued.

Host phenotype		Mammalian host species		Links between variations of the microbiome and the host phenotype	References
Energy uptake from diet	Seasonal diet shifts	Lemurs	*Lemur catta and Propithecus verreauxi*	Seasonal shifts in diet were associated with the shifts in microbiota.	Fogel et al. 2015
		Wood mouse	*Apodemus sylvaticus*		Maurice et al. 2015
		Reindeer*	*Raingifer tarandus tarandus*		Salgado-Flores et al. 2016
		Gorrila	*Gorilla gorilla sp.*	Seasonal shifts in diet were associated with shifts in the microbiota and the metabolome.	Gomez et al. 2016
		Black Howler Monkey	*Alouatta pigra*	Seasonal shifts in diet were associated with shifts in the microbiota and volatile fatty acids.	Amato et al. 2014
		Tibetan Macaques	*Macaca Thibetana*	Seasonal shifts in diet were associated with shifts in microbiota and predicted bacterial gene functions.	Sun et al. 2016
	Mucin digestion during food limitation	Thirteen-lined ground squirrel	*Ictidomys tridecemlineatus*	Fasting was associated with changes in the microbiota. The increase in acetate (%) was associated with the increase in acetate producing bacteria (%) during fasting and hibernation.	Carey et al. 2013
		Arctic ground squirrel	*Urocitellus parryii*		Stevenson et al. 2014
		Syrian hamster*	*Mesocricetus auratus*	Fasting was associated with changes in the microbiota and the increases in acetate producing bacteria.	Sonoyama et al. 2009
Detoxification of plant dietary toxins	Oxalate detoxification	White-throated woodrat	*Neotoma albigula*	Bacterial communities that are associated with oxalate-degradation are most abundant in the foregut compared to the rest of the gastrointestinal tract.	Miller et al. 2014
	Resin detoxification	Desert woodrat	*Neotoma lepida*	Population differences in microbial composition and functions were associated with creosote diet. Antibiotic treatment and transplant experiments demonstrated a causal link.	Kohl et al. 2014b
	Tannin detoxification			Transfering the microbiome of woodrats to a lab rats revlealed that differences in the microbial community structure can increase the ability of recipients to comsume tannic acid.	Kohl et al. 2016b

* Captive individuals in semi-natural environment were used mostly or entirely.

Table 1. Continued.

Host phenotype		Mammalian host species		Links between variations of the microbiome and the host phenotype	References
Immunity and pathogen defense	Interaction with viruses	Humans	*Homo sepiens*	Immunodeficiency due to viral infection was associated with shifts in the microbiota.	reviewed in Salas and Chang 2014, Williams et al. 2016
		Chimpanzee	*Pan troglodytes*		Moeller et al. 2013b
		Gorilla	*Gorilla gorila gorilla*		Moeller et al. 2015
	Interaction with macroparasites	House mice	*Mus musculus domesticus*	Antibody-based viral infection status was positively correlated with alpha-diversity of microbiota.	Weldon et al. 2015
				Infection statuses of nematodes and mites were negatively correlated with alpha-diversity of the microbiota.	
		Yellow-necked mouse	*Apodemus flavicollis*	A weak association was observed between microbial composition and helminth composition.	Kreisinger et al. 2015
	Immune gene expression	House mice	*Mus musculus spp.*	Aberrant immune gene expression and intestinal histopathology in hybrids were associated with differences in the microbiota in relation to the parental species.	Wang et al. 2015
Chemical communication and behavior	Olfactory signaling in scent glands	Spotted and Striped hyenas	*Crocuta crocuta* and *Hyaena hyaena*	The microbiota including odor-producing bacteria in scent gland secretions differed among species, social groups, sex, and reproductive status using pyrosequencing.	Theis et al. 2012, 2013
		Meerkat	*Suricata suricatta*	The microbiota (based on ARISA) in scent secretion differed among sex, age, and social groups.	Leclaire et al. 2014
		Musk deer*	*Moschus berezovskii*	The chemical and microbial composition of musk secretion differed between mated and unmated males using pyrosequencing.	Li et al. 2016
		European Badger	*Meles meles*	The chemical and microbial composition (based on T-RFLP) of subcaudal gland secretion were significantly correlated and varied by age and female reproductive status.	Sin et al. 2012

* Captive individuals in semi-natural environment were used mostly or entirely.

1.4. Chapter 1 Figure

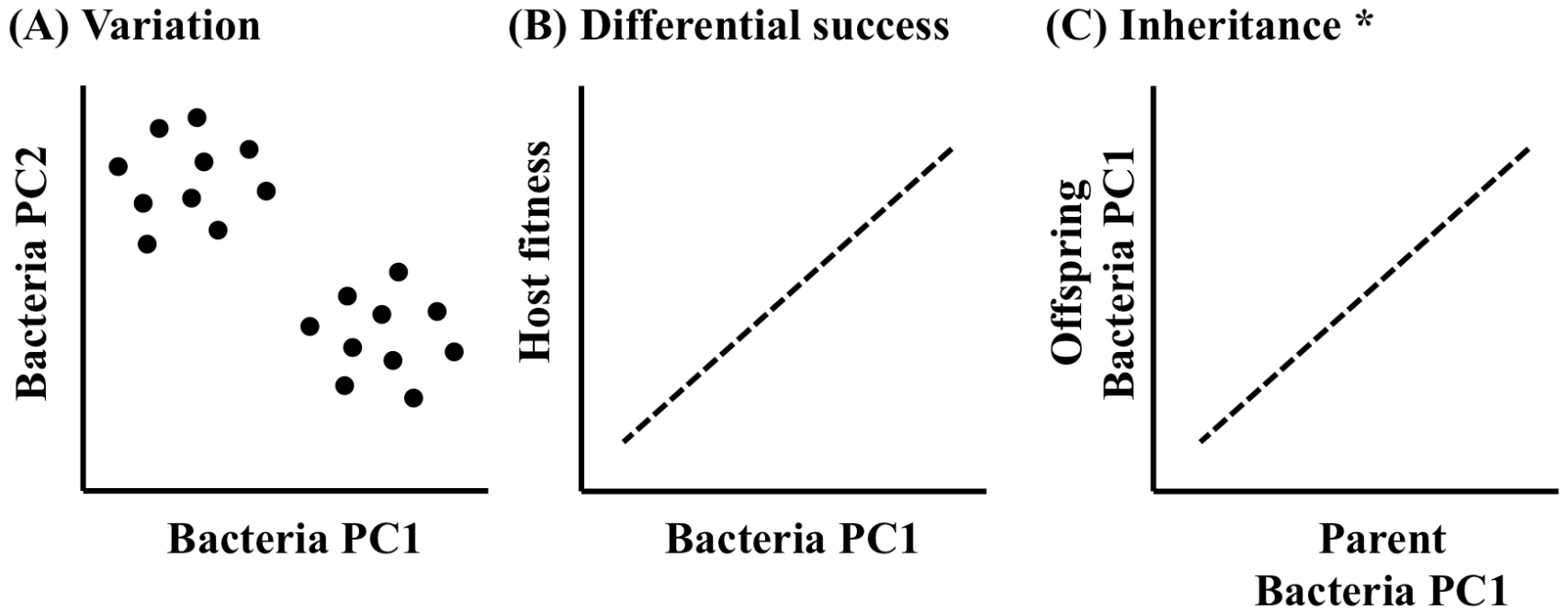


Figure 1. Three components of natural selection applied to the host-associated microbiome. (A) Evidence of variation in microbiome represented by hypothetical PCA plot of microbial taxa or genes. (B) Evidence of the differential success of the host based on the differences in the microbiome. (C) Evidence of inheritance of the microbiome, where offspring microbiome resembles the parent microbiome. All three components are required for the host to evolve in response to the natural selection on the microbiome across multiple generations. * See the text for genetic and non-genetic mechanisms of microbiome inheritance.

Chapter 2

Geographical variation of gut microbial composition in humans

This chapter has been previously published and is reproduced here with permission from the co-author and in accordance with the journal's article sharing policy:

Suzuki, T.A. and M. Worobey. 2014. Geographical variation of gut microbial composition in humans. Biology Letters. 10(2): 20131037.

Abstract

Although we know there is considerable variation in gut microbial composition within host species, little is known about how this variation is shaped and why such variation exists. In humans, obesity is associated with the relative abundance of two dominant bacterial phyla: an increase in the proportion of Firmicutes and a decrease in the proportion of Bacteroidetes. Since there is evidence that humans have adapted to colder climates by increasing their body mass (e.g. Bergmann's rule), we tested whether Firmicutes increase and Bacteroidetes decrease with latitude, using 1020 healthy individuals drawn from 23 populations and six published studies. We found a positive correlation between Firmicutes and latitude and a negative correlation between Bacteroidetes and latitude. The overall pattern appears robust to sex, age, and bacterial detection methods. Comparisons between African Americans and native Africans and between European Americans and native Europeans suggest no evidence of host genotype explaining the observed patterns. The variation of gut microbial composition described here is consistent with the pattern expected by Bergmann's rule. This surprising link between large-scale geography and human gut microbial composition merits further investigation.

2.1. Introduction

There is growing appreciation that the gut microbial community (i.e. microbiota) may have played an important role in the evolution of host species (Muegge *et al.* 2011). Although the gut microbial composition is quite distinctive between host species (Muegge *et al.* 2011), there is also considerable variation within host species (Lay *et al.* 2005; Mueller *et al.* 2006; Turnbaugh *et al.* 2009; De Filippo *et al.* 2010; Wang *et al.* 2012; Yatsunenko *et al.* 2012). Nonetheless, how within-species microbial variation is shaped and why such variation exists remains largely unexplored.

One of the factors that is associated with gut microbial composition is host physiology (Bäckhed *et al.* 2004; Ley *et al.* 2005, 2006; Jumpertz *et al.* 2011). For example, obese mice have distinct bacterial composition compared to lean mice characterized by relative abundance of two dominant phyla: increase in Firmicutes and decrease in Bacteroidetes (Ley *et al.* 2005). Lean mice transplanted with the microbiota

from obese mice demonstrated increased fat storage without increased food consumption when compared to control mice (Bäckhed *et al.* 2004). This suggests that relative increases in Firmicutes and decreases in Bacteroidetes can increase energy extraction and fat storage from a given diet. Interestingly, the pattern is consistent in humans where increases in Firmicutes and decreases in Bacteroidetes are associated with obesity (Ley *et al.* 2006) and also associated with increased energy harvest from the diet (Jumpertz *et al.* 2011).

It seems plausible that an increase in energy extraction and fat storage may be more important for animals in colder regions compared to animals in warmer regions as an environmental adaptation to climate. It is generally accepted that humans follow 'Bergmann's rule' (Bergmann 1847) at least to some degree, whereby populations in higher latitudes tend to have larger body mass compared to populations in lower latitudes (Roberts 1953). We speculated that such large-scale eco-geographic observations might be mediated in part by gut microbial communities that modulate energy extraction and fat storage. However, few studies have applied ecological theories to the geographic variation of gut microbial composition in humans or other animals across broad latitudinal scales. Here, we test whether the relative of abundance of Firmicutes in the human gut increases with latitude and whether that of Bacteroidetes decreases (i.e. the pattern that might be expected given body mass variation with latitude) using six published human microbial studies (Lay *et al.* 2005; Mueller *et al.* 2006; Turnbaugh *et al.* 2009; De Filippo *et al.* 2010; Wang *et al.* 2012; Yatsunenko *et al.* 2012), including more than twenty populations from a variety of geographic locations worldwide.

2.2. Methods

We searched human gut microbial studies that included population samples. We either requested the data for each individual used in the study from the original authors or acquired the data from MG-RAST (Meyer *et al.* 2008). We were able to access six microbial studies in this way (Lay *et al.* 2005; Mueller *et al.* 2006; Turnbaugh *et al.* 2009; De Filippo *et al.* 2010; Wang *et al.* 2012; Yatsunenko *et al.* 2012) including 23 population samples representing a total of 1020 healthy individuals (Table 1, S8, Fig. 1). The relative abundance of Firmicutes (i.e. Firmicutes/total bacteria), the relative abundance of Bacteroidetes (i.e. Bacteroidetes/total bacteria), sex, age, sampling localities, and bacterial detection methods were collected from the studies. Age was divided into five age classes: Z (<1), A (1-10), B (11-20), C (21-54), and D (>60). Latitudes were based on sampling locality of the population. The bacterial detection methods were divided into two categories: FISH-based or 16S-sequencing-based methods. Ethnicity information was also collected when available. Spearman correlations were used for all correlations and the Wilcoxon rank sum test was used for all pairwise comparisons.

2.3. Results and discussion

There was a highly significant positive correlation between Firmicutes abundance and latitude (rho = 0.857, $P < 0.0001$) and a negative correlation between Bacteroidetes

and latitude (rho = -0.637, $P = 0.001$) using the population averages from all 23 population samples (Fig. S1). The pattern is robust even when considering individual data points (Table 2 and Fig. 2). Although one might expect opposite correlations between the two phyla since the majority of the human gut microbiome consists of Firmicutes and Bacteroidetes (i.e. if one phylum increases the other phylum is likely to decrease), the correlation between Firmicutes and latitude is consistently greater compared to Bacteroidetes and latitude. This pattern can be explained either by taxa in Firmicutes mainly driving the pattern relative to Bacteroidetes or by taxa in minor phyla (e.g. Proteobacteria, Actinobacteria, Tenericutes, etc.) sharing some functional roles with Bacteroidetes. How individual bacterial taxa relate to the net effect of the microbiota remains an open question. Regardless, the correlations observed here are consistent with the pattern expected by Bergmann's rule (Roberts 1953) where increases in Firmicutes and decreases in Bacteroidetes are known to be associated with an increase in body weight (Ley *et al.* 2005) potentially due to an increase in energy extraction and fat storage from a given diet (Bäckhed *et al.* 2004; Jumpertz *et al.* 2011). However, the pattern might be explained not only by latitude but also perhaps by age, sex, or bacterial detection methods. Therefore, we tested the potential factors that might be driving the observed pattern as described below.

First, it is unlikely that age is biasing the overall pattern. Although bacterial composition differed among different age classes (Table S1), all of the significant correlations within all age classes were in the directions expected by Bergmann's rule except in the >60 age class (Table S2). Interestingly, elderly individuals showed significant correlations in the opposite direction for both bacterial phyla (Table S2). However, elderly individuals (and infants) are known to have less stable and distinct gut microbial compositions compared to non-elderly adults (Spor, Koren and Ley 2011); we therefore analyzed the data using not only all of the data (n=1020) but also a subset of data representing non-elderly adults (age class C, n=438) (Table 2), with similar results in each case.

Second, sex is apparently not biasing the results since the relative abundances of two bacterial phyla between men and women are similar (Table S3) and the correlation between the bacterial phyla and latitude remained significant within each sex (Table S4).

Next, we tested potential biases between different microbial detection methods. The populations with FISH-based methods had higher Firmicutes and lower Bacteroidetes values compared to populations investigated with 16S-sequencing-based methods (Table S5). This could cause a potential bias in our data set since FISH-based methods were concentrated in the higher latitudinal populations and 16S-sequencing-based methods were concentrated in the lower latitudinal populations (Table 1). However, there were significant correlations within populations with 16S-sequencing-based methods alone (Table 2 and S7). Although the Spearman rho values became weaker, the overall pattern cannot be explained by the potential method bias alone. This issue could in the future be resolved by characterizing gut microbial composition in European samples using consistent 16S-sequencing-based methods. Sampling an independent latitudinal transect in the Southern hemisphere would also help to test this pattern in a more robust way.

Finally, to investigate whether there is an effect of host genotype on the variation of gut microbiota, we focused on four ethnic groups: native Africans (AF), native

Europeans (EU), African Americans (AA), and European Americans (EA) (Turnbaugh *et al.* 2009; De Filippo *et al.* 2010; Yatsunenko *et al.* 2012). All populations were studied using 16S-sequencing methods. Applying the concept of a common garden experiment, if there is a host genotype effect on gut microbial composition, we expect populations of African ancestry (i.e. AF and AA) to have similar composition to each other compared to populations of European ancestry (i.e. EU and EA) due to genetic relatedness. The result did not support the prediction of host genotype effect: abundance of bacterial phyla was always different in native African populations (AF) from the other populations (i.e. EU, AA, and EA) (Table S7). Although quantitative measures of diet were unavailable from most populations, the pattern could potentially be explained by the effect of diet since at least AF-Burkina Faso had a low fat/high fiber diet and EU, AA, and EA had high fat/low fiber diet or non-restricted typical western diet (Turnbaugh *et al.* 2009; De Filippo *et al.* 2010). Also, climate and pathogen/parasite distribution might potentially explain the overall pattern. Dissecting the co-varying factors of latitude may help identify the potential driver of gut microbial variation.

If gut microbial community is playing a role in Bergmann's rule (or, more generally, size or physiology differences between latitudes), there are three major models that could explain the observed pattern, and they are not mutually exclusive. I) Host genes could maintain the observed microbial community variation as an environmental adaptation. For example, immune genes such as antimicrobial peptides or behavioral genes that can alter food preferences or amount of food intake could potentially regulate the microbiota. II) The environment itself could maintain the observed microbiota variation if a key environmental factor is co-varying with latitude. For example, if food availability (e.g. food cultural variation in the case of humans) or microbial availability were correlated with latitude, environmental variation itself could explain the pattern without a fitness consequence of microbes to the host. III) Host plasticity could also maintain the observed microbiota variation. For example, cold environment increases food intake in animals (Brobeck 1948). If different microbiota have different fitness consequence to the host (e.g. different microbiota add different amounts of fat), non-genetic vertical or horizontal transmission of microbes could maintain the observed pattern. Which of the models explains the observed variation of human gut microbiota remains an open question.

Although further investigation is necessary to characterize the geographic variation of gut microbial composition in humans, the robust pattern raises some interesting points in microbial ecology. 1) "Healthy microbiota" may differ in different geographic regions. 2) Independent latitudinal transects using consistent methods may help to identify environmental variables that shape gut microbial composition. 3) Gut microbial composition could potentially help mediate the fit of host phenotype to its environment with or without a host genotype effect. 4) We can generate novel hypotheses by understanding the geographic variation of host-associated microbial communities within species. Studying other species that show clinal variation in body size may help establish whether the observations reported here are general.

2.4. Chapter 2 Tables

Table 1. Six studies used in this study

sample locations	latitude	sample size	# of individuals per age class						female (%)	reference	map ID[1]
			Z (<1)	A (1-10)	B (11-20)	C (21-54)	D (>60)	No data			
Coromoto, Venezuela	5.4	53	6	19	6	16	4	2	57	7	1
Platanillal, Venezuela	5.4	45	1	24	7	8	5	-	58	7	2
Burkina Faso	12.3	14	-	14	-	-	-	-	36	2	3
Chamba, Malawi	15.3	21	5	6	-	3	-	7	56	7	4
Mayaka, Malawi	15.4	51	9	17	-	12	-	13	73	7	5
Mbiza, Malawi	16.0	31	11	5	-	6	-	9	83	7	6
Makwhira, Malawi	16.2	5	2	1	-	1	-	1	75	7	7
Shanghai, China	31.2	56	-	-	-	56	-	-	54	6	8
Missouri, USA	38.6	120	-	-	82	38	-	-	100	5	9
Missouri, USA	38.6	30	-	-	-	23	-	7	100	7	10
St.Louis, USA	38.6	235	22	12	106	95	-	-	54	7	11
Boulder, USA	40.0	9	1	7	-	1	-	-	67	7	12
Philadelphia, USA	40.0	23	-	13	-	10	-	-	78	7	13
Camerino, Italy	43.1	58	-	-	-	20	38	-	53	4	14
Tuscany, Italy	43.4	13	-	13	-	-	-	-	31	2	15
Paris, France	48.9	48	-	-	-	22	26	-	58	4	16
France	48.9	21	-	-	-	21	-	-	52	3	17
United Kingdom	51.5	10	-	-	-	10	-	-	80	3	18
Netherlands	52.4	20	-	-	-	20	-	-	70	3	19
Potsdam, Germany	52.4	58	-	-	-	21	37	-	62	4	20
Germany	52.5	20	-	-	1	19	-	-	80	3	21
Denmark	55.7	20	-	3	1	16	-	-	55	3	22
Stockholm, Sweden	59.3	59	-	-	-	20	39	-	53	4	23

Table2. Correlations with latitude

phyla	data	n	correlations between latitude	
			Spearman's ρ	p-value
Firmicutes (%)	whole data	1020	0.563	<0.0001
	whole data (age C)	438	0.606	<0.0001
	only 16S seq. data	706	0.434	<0.0001
	only 16S seq. data (age C)	269	0.375	<0.0001
Bacteroidetes (%)	whole data	1020	-0.279	<0.0001
	whole data (age C)	438	-0.464	<0.0001
	only 16S seq. data	706	-0.117	0.002
	only 16S seq. data (age C)	269	-0.319	<0.0001

2.5. Chapter 2 Figures

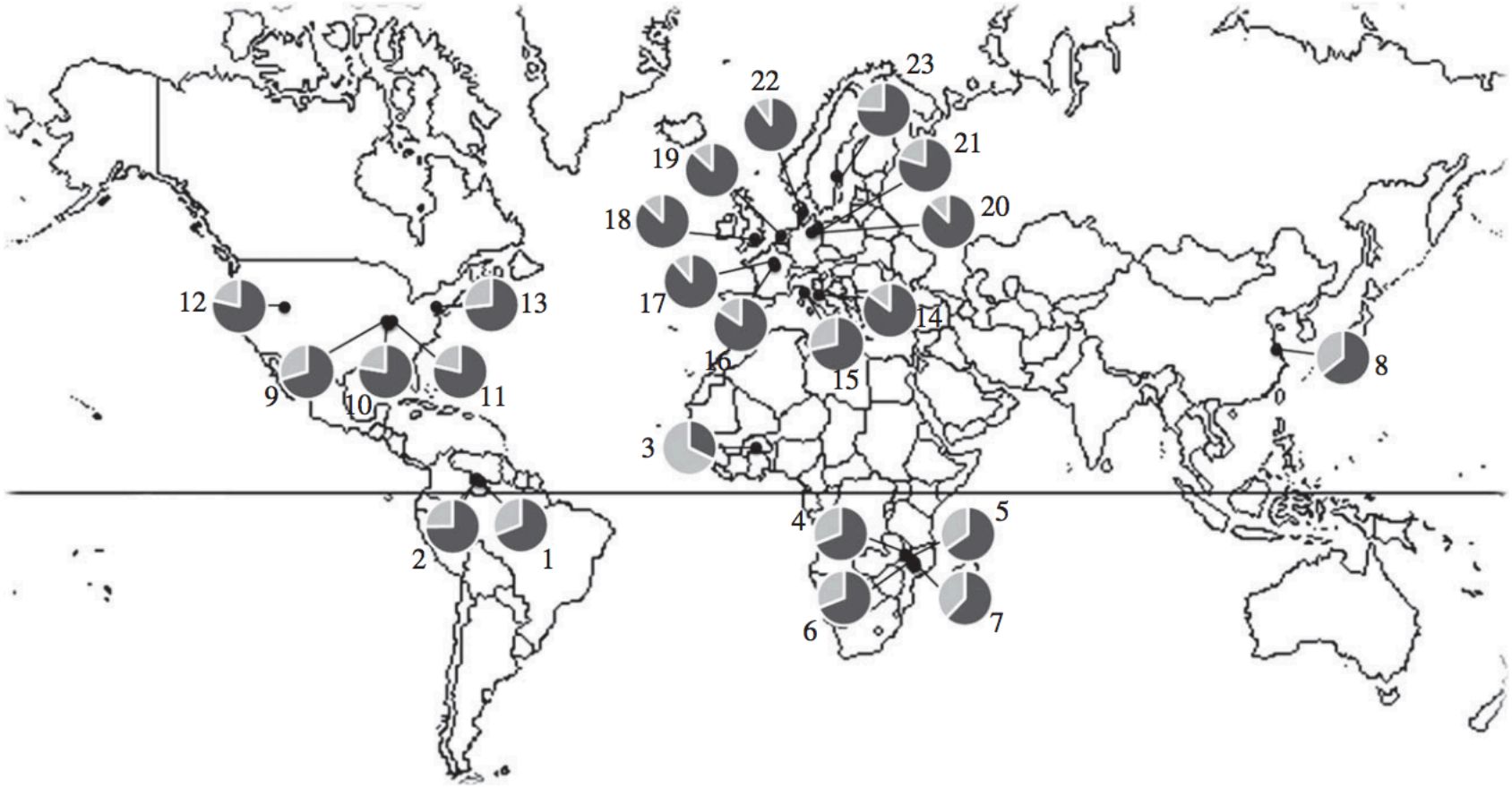


Figure 1. Distribution map of human populations used in this study. Pie chart indicates the relative abundance of two bacterial phyla: Firmicutes (Light gray) and Bacteroidetes (Dark gray). The numbers indicate Map ID (see Table 1).

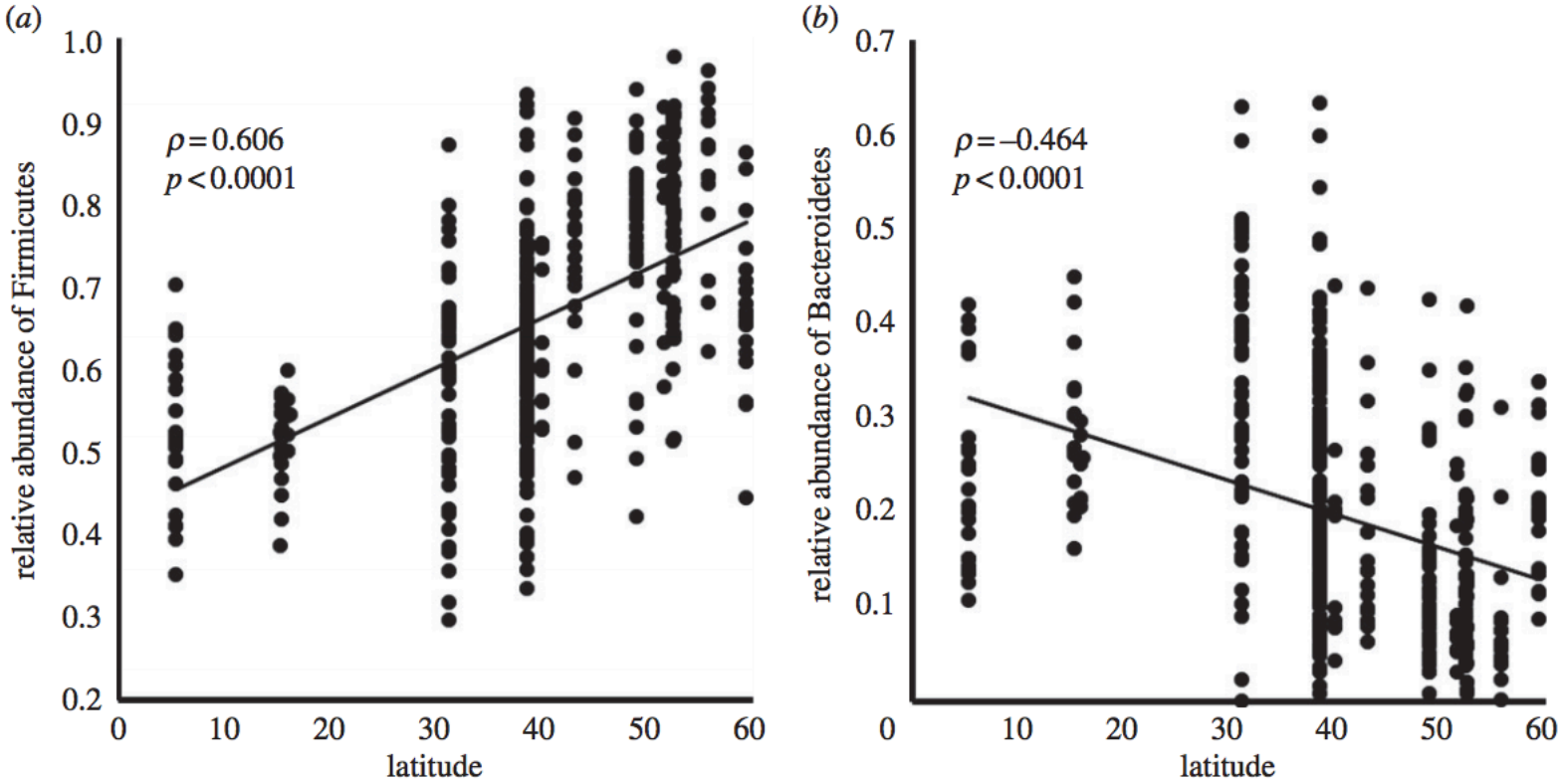


Figure 2. Correlations between (a) relative abundance of Firmicutes and latitude and between (b) relative abundance of Bacteroidetes and latitude for all adults (n = 438).

2.6. Chapter 2 Supplemental information

Table S1. Effects of age (Wilcoxon rank sum test)

phyla	data	n	age classes	mean	s.d.	significance[1]			
Firmicutes	whole data	57	Z (0-1)	0.42	0.16	a			
		134	A (1-10)	0.52	0.15		b		
		203	B (11-19)	0.64	0.12			c	
		438	C (20-54)	0.66	0.14			c	
		149	D (>60)	0.70	0.14				d
Bacteroidetes	whole data	57	Z (0-1)	0.15	0.12	a			
		134	A (1-10)	0.25	0.17				d
		203	B (11-19)	0.22	0.13			c	d
		438	C (20-54)	0.20	0.12		b	c	
		149	D (>60)	0.17	0.11	a	b		

[1] Rows not connected by same letter are significantly different, Bonferroni corrected α=0.005

Table S2. Correlations within age class (Spearman rho)

phyla	data	n	age classes	correlations between latitude	
				Spearman rho	*p* value
Firmicutes	whole data	57	Z (0-1)	0.342	**0.009**
		134	A (1-10)	0.361	**<0.0001**
		203	B (11-19)	0.200	**0.004**
		438	C (20-54)	0.606	**<0.0001**
		149	D (>60)	-0.248	**0.002**
Bacteroidetes	whole data	57	Z (0-1)	0.000	0.99
		134	A (1-10)	0.080	0.4
		203	B (11-19)	-0.091	0.2
		438	C (20-54)	-0.464	**<0.0001**
		149	D (>60)	0.334	**<0.0001**

TableS3. Effects of sex (Wilcoxon rank sum test)

phyla	data	n	sex	mean	s.d.	*p* value
Firmicutes	whole data	636	Female	0.63	0.15	0.8
		349	Male	0.63	0.15	
	whole data (age C)	308	Female	0.65	0.14	**0.02**
		128	Male	0.68	0.13	
Bacteroidetes	whole data	636	Female	0.21	0.13	**0.04**
		349	Male	0.19	0.13	
	whole data (age C)	308	Female	0.21	0.13	0.1
		128	Male	0.18	0.12	

Table S4. Correlations within sex (Spearman rho)

phyla	data	n	sex	correlations between latitude	
				Spearman rho	*p* value
Firmicutes	whole data	637	Females	0.491	**<0.0001**
		349	Males	0.474	**<0.0001**
	whole data (age C)	308	Females	0.638	**<0.0001**
		128	Males	0.508	**<0.0001**
Bacteroidetes	whole data	637	Females	-0.313	**<0.0001**
		349	Males	-0.312	**<0.0001**
	whole data (age C)	308	Females	-0.480	**<0.0001**
		128	Males	-0.420	**<0.0001**

Table S5. Effects of Method (Wilcoxon rank sum test)

phyla	data	n	methods	mean	s.d.	*p* value
Firmicutes	whole data	706	16S seq.	0.58	0.14	**<0.0001**
		314	FISH	0.73	0.12	
	whole data (age C)	269	16S seq.	0.60	0.11	**<0.0001**
		169	FISH	0.76	0.11	
Bacteroidetes	whole data	706	16S seq.	0.23	0.13	**<0.0001**
		314	FISH	0.15	0.1	
	whole data (age C)	269	16S seq.	0.24	0.13	**<0.0001**
		169	FISH	0.14	0.09	

Table S6. Correlations within methods (Spearman rho)

phyla	data	n	methods	correlations between latitude	
				Spearman rho	*p* value
Firmicutes	whole data	706	16S Seq.	0.434	**<0.0001**
	Only adults (age C)	269	16S Seq.	0.375	**<0.0001**
	whole data	314	FISH	-0.244	**<0.0001**
	Only adults (age C)	169	FISH	-0.028	0.7
Bacteroidetes	whole data	706	16S Seq.	-0.117	**0.002**
	Only adults (age C)	269	16S Seq.	-0.319	**<0.0001**
	whole data	314	FISH	0.201	**0.0003**
	Only adults (age C)	169	FISH	-0.013	0.9

Table S7. Effects of ancestry (Wilcoxon rank sum test)

phyla	data	n	ethinc groups	mean	s.d.	significance[1]		
Firmicutes	whole data	14	AF (Brkina Faso)	0.28	0.17	a		
		108	AF (Malawi)	0.48	0.17		b	
		81*	AF (Malawi)*	0.51*	0.08*		b*	
		13	EU	0.65	0.15			c
		62	AA	0.66	0.15			c
		58	EA	0.69	0.13			c
Bacteroidetes	whole data	14	AF (Brkina Faso)	0.58	0.26	a		
		108	AF (Malawi)	0.24	0.11		b	
		81*	AF (Malawi)*	0.26*	0.08*		b*	
		13	EU	0.26	0.09		b	
		62	AA	0.29	0.15		b	
		58	EA	0.28	0.13		b	

* age class Z (<1) excluded

[1] Rows not connected by same letter are significantly different, Bonferroni corrected α=0.005

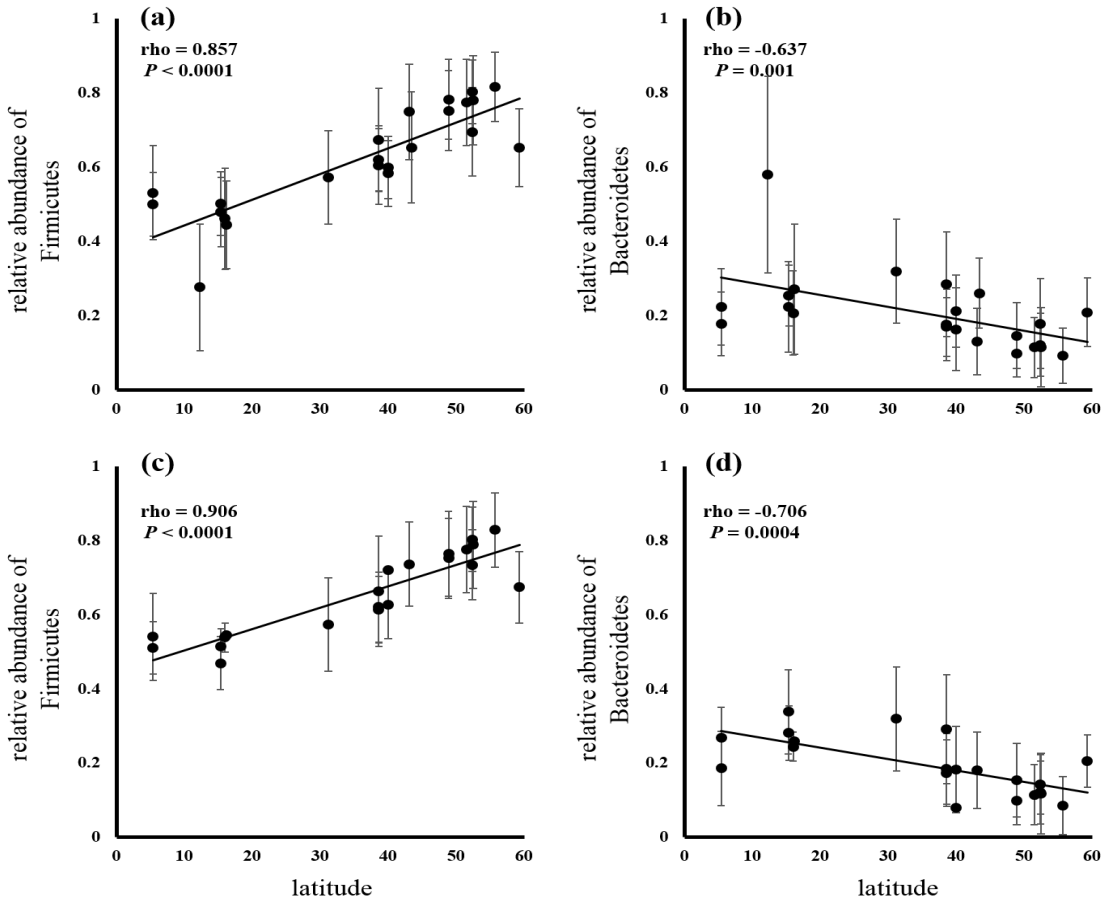


Figure S1. Correlation between latitude and population average bacterial abundance. The error bar shows standard deviations. (a) Correlation between latitude and average relative abundances of Firmicutes representing 23 populations (all data, n=1020). (b) Correlation between latitude and average relative abundances of Bacteroidetes representing 23 populations (all data, n=1020). (c) Correlation between latitude and average relative abundances of Firmicutes representing 21 populations (only adults, n=438). (d) Correlation between latitude and average relative abundances of Bacteroidetes representing 21 populations (only adults, n=438).

Chapter 3

Spatial heterogeneity of gut microbial composition along the gastrointestinal tract in natural populations of house mice.

This chapter has been previously published and is reproduced here with permission from the co-author and in accordance with the journal's article sharing policy:

Suzuki, T.A. and M.W. Nachman. 2016. Spatial heterogeneity of gut microbial composition along the gastrointestinal tract in natural populations of house mice. PLOS ONE 11(9): e0163720.

Abstract

There is a growing appreciation of the role of gut microbial communities in host biology. However, the nature of variation in microbial communities among different segments of the gastrointestinal (GI) tract is not well understood. Here, we describe microbial communities from ten different segments of the GI tract (mouth, esophagus, stomach, duodenum, ileum, proximal cecum, distal cecum, colon, rectum, and feces) in wild house mice using 16S rRNA gene amplicon sequencing. We also measured carbon and nitrogen stable isotopic ratios from hair samples of individual mice as a proxy for diet. We identified factors that may explain differences in microbial composition among gut segments, and we tested for differences among individual mice in the composition of the microbiota. Consistent with previous studies, the lower GI tract was characterized by a greater relative abundance of anaerobic bacteria and greater microbial diversity relative to the upper GI tract. The upper and lower GI tracts also differed in the relative abundances of predicted microbial gene functions, including those involved in metabolic pathways. However, when the upper and lower GI tracts were considered separately, gut microbial composition was associated with individual mice. Finally, microbial communities derived from fecal samples were similar to those derived from the lower GI tract of their respective hosts, supporting the utility of fecal sampling for studying the gut microbiota of mice. These results show that while there is substantial heterogeneity among segments of the GI tract, individual hosts play a significant role in structuring microbial communities within particular segments of the GI tract.

3.1. Introduction

Recent advances in microbial ecology have demonstrated important roles of gut microbes in digestion (Hooper, Midtvedt and Gordon 2002), immunity (Round and Mazmanian 2009b), development (Fraune and Bosch 2010), and behavior (Heijtz *et al.* 2011) of hosts. Despite the importance of gut microbial communities in host biology, many studies depend solely on fecal samples to investigate the gut microbial community.

Less attention has been given to spatial heterogeneity along the gastrointestinal (GI) tract or to the mechanisms structuring microbial variation along the GI tract.

In mammals, diet seems to be a major driver of fecal microbial communities within (Carmody *et al.* 2015) and between species (Ley *et al.* 2008a; Muegge *et al.* 2011). However, hindgut and foregut fermenters show distinct fecal microbial communities despite eating similar herbivorous diets (Ley *et al.* 2008a; Muegge *et al.* 2011). Hindgut fermenters have a simple stomach, and the majority of fermentation takes place in the enlarged cecum. In contrast, foregut fermenters have, in addition to the cecum, a segmented stomach where the majority of fermentation takes place. This suggests that fecal microbial communities could be partly determined by gut anatomy (Ley *et al.* 2008a) along with other factors including host genotype (Spor, Koren and Ley 2011) and geography (Suzuki and Worobey 2014).

One approach to understanding the factors affecting the gut microbiota along the GI tract is to assess the relative importance of gut segments and individual hosts in explaining the composition of the microbial community. For example, if variation among gut segments explains most of the variation in microbial composition, this might suggest that gut-segment-specific biochemical differences (e.g. pH, oxygen, nutrients, etc.) determine microbial communities along the GI tract (Durbán *et al.* 2011; Stearns *et al.* 2011; Gu *et al.* 2013; Kohl *et al.* 2014a; Lu *et al.* 2014). In contrast, if variation among individuals explains most of the microbial variation, this would suggest that differences in diet, host genotype, and/or host habitat are important in determining the composition of microbial communities (Eckburg *et al.* 2005; Zhang *et al.* 2014). The two hypotheses above are not mutually exclusive.

Characterizing the complete GI tract is important not only for understanding microbial composition in different gut segments but also for assessing the efficacy of fecal sampling for studying gut microbial communities. The assumption that fecal microbial communities are similar to those in the colon has been examined mostly in humans, and the two types of samples usually show significant differences (Ott *et al.* 2004; Ouwehand *et al.* 2004; Eckburg *et al.* 2005; Durbán *et al.* 2011; Stearns *et al.* 2011). Recently, multiple sites along the GI tract have also been characterized in laboratory mice (Gu *et al.* 2013), wood rats (Kohl *et al.* 2014a), flying squirrels (Lu *et al.* 2014), and rhesus macaques (Yasuda *et al.* 2015). For example, in wood rats, the fecal microbial community is different from that in the large intestine and is more similar to the communities found in the stomach or small intestine (Kohl *et al.* 2014a). In contrast, in rhesus macaques, the relative abundances of bacterial taxa in fecal samples were significantly correlated with those in both the large and small intestines (Yasuda *et al.* 2015). Together, these studies suggest that fecal microbial communities may be most similar to different parts of the GI tract in different host species.

Wild house mice provide a useful mammalian model for studying microbial variation along the GI tract. Laboratory strains of house mice have been used extensively as a model in gut microbial ecology. Microbial variation along the GI tract has been characterized in one strain of inbred mice in a laboratory environment (Gu *et al.* 2013). The authors sampled seven locations along the GI tract in six individuals of a single genotype, and found differences in the microbial community mainly between the upper and lower GI tract (Gu *et al.* 2013). Wild mice are genetically variable, and host genotype is known to play an important role in shaping the gut microbiota (Benson *et al.* 2010b).

Wild mice may also show variation in microbial communities due to differences among animals sampled from different localities (Linnenbrink *et al.* 2013; Weldon *et al.* 2015). In addition, laboratory diets and environment have been shown to alter the gut micobiota in *Drosophila* (Chandler *et al.* 2011) and house mice (Kreisinger *et al.* 2014; Wang *et al.* 2014, 2015) compared to their wild relatives. Finally, house mice live in close association with humans. The parallel ecology between humans and house mice highlights the importance of studying the microbial composition of wild house mice for translational research (Guénet and Bonhomme 2003; Phifer-Rixey and Nachman 2015).

Here, we collected wild house mice in Tucson, Arizona. First we describe the microbial heterogeneity along ten segments of the GI tract. We then assess the relative importance of gut segment, host individual, and diet in explaining the composition of the microbial communities observed in our samples.

3.2. Methods

3.2.1. Animal and sample collections

Three adult male and three adult female *Mus musculus* were collected from Tucson, Arizona on September 19, 2012 (Table S1 in Supplemental information). Animals were collected on private property with the permission of the landowners under a State of Arizona scientific collecting permit (LIC# SP791101 to Taichi Suzuki). Sherman live traps were used without bait to assess the "natural" gut microbial composition. Animals were collected from four localities (Table S1 in Supplemental information). Animals were kept in Sherman traps, euthanized by cervical dislocation, and all samples were collected within 24 hours after capture. All procedures involving animals were reviewed and approved by the University of Arizona Institutional Animal Care and Use Committee (protocol 07-004). Museum specimens (skins and skulls) were prepared and have been deposited in the mammal collection of the Museum of Vertebrate Zoology at the University of California, Berkeley (catalog numbers MVZ230525, 230526, 230530, 230535, 230537, and 230539).

The samples were collected from the following 10 locations along the GI tract under sterile conditions: oral cavity, esophagus, stomach, duodenum, ileum, proximal half of cecum, distal half of cecum, proximal colon, rectum, and fresh feces (Fig 1A). The oral cavity was swabbed using 1 cm^2 kimwipe. All tissues were collected immediately after the animal was euthanized. All samples were stored immediately at -80°C.

3.2.2. Stable isotope analyses

Hair samples were used to infer diet based on carbon and nitrogen stable isotope ratios. Hair samples (0.4 mg) were collected from the base of the right hind leg from each individual. The hair was rinsed in 2:1 chloroform:methanol to remove the surface oils, rinsed in distilled water, and dried in collection tubes. Stable isotope analyses were performed in the Environmental Isotope Laboratory at the University of Arizona. Carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) stable isotope ratios were measured on a continuous-flow gas-

ratio mass spectrometer (Finnigan Delta PlusXL), which was coupled to an elemental analyzer (Costech). Standardization was based on acetanilide for elemental concentration ($\delta^{13}C$: NBS-22 and USGS-24, $\delta^{15}N$: IAEA-N-1 and IAEA-N-2).

3.2.3. DNA extraction and Sequencing

The frozen samples were chopped into pieces with a sterile razor blade in a petri dish on dry ice. We used 200 mg of each sample, except for those from the mouth and esophagus which were smaller. To study microbes living in the gut contents and on the mucosal surface, the entire sample with the host tissue was immediately disrupted and rinsed with sterile forceps in 1.4ml ASL (from QIAamp DNA stool Minikit). Most of the host tissue was removed by this procedure. Following a protocol developed by Smith et al. (2011) to extract microbial DNA (Smith *et al.* 2011), 0.2 g of sterile zirconia/silica disruption beads (0.1mm, Research Products International Corp.) were added and vortexed until the samples were thoroughly homogenized. Samples were then placed in a TissueLyser LT (Qiagen) 30HZ for 6 min for further mechanical disruption. The suspension was heated at 95°C for 5 min, and DNA was extracted following step 4 (vortex and centrifugation) in the protocol from the QIAamp DNA stool Minikit (Qiagen). The quality of the DNA was evaluated based on Nano Drop 3300 (Thermo Scientific). DNA samples were stored at -20°C before sequencing.

The DNA samples were shipped to Argonne National Laboratory for 16S rRNA amplicon sequencing at their Next Generation Sequencing Core Facility. The V4 region of the 16S rRNA gene was amplified and the samples were multiplexed for sequencing on a 150bp paired-end Illumina MiSeq platform using primers 515F (5' - GTGCCAGCMGCCGCGGTAA) 806R (5' - GGACTACHVGGGTWTCTAAT) and barcodes described in (Caporaso *et al.* 2010). All data have been deposited into the European Nucleotide Archive (accession number PRJEB15238).

3.2.4. Analyses

The paired-end reads were merged in PANDAseq (Masella *et al.* 2012) and all the merged sequences were analyzed in QIIME version 1.8.0 (Caporaso *et al.* 2010). Sequences were demultiplexed and quality-filtered using the default parameters in QIIME. Chimeric sequences were detected and removed by following both the reference (Greengenes 13_8) (DeSantis *et al.* 2006) and de-novo based approaches using USEARCH/UCHIME 6.1 (Edgar 2010; Edgar *et al.* 2011). One sample (TAS203.7, Table S1 in Supplemental information) was excluded from all analyses due to low sequence reads. We assigned sequences to OTUs by using the subsampling open-reference approach with default parameters in QIIME (Rideout *et al.* 2014). Briefly, sequences were initially clustered against a reference database (Greengenes 13_8) (DeSantis *et al.* 2006) using UCLUST (Edgar 2010) with a minimum sequence identity of 97%. A random subsample of the reads that fail to hit the reference database was subsequently clustered de novo with default parameters. Singleton OTUs were removed. Taxonomy was assigned using uclust-based consensus taxonomy with default parameters. A phylogenetic tree was created from representatives of all OTUs using FastTree (Price, Dehal and Arkin 2009). The OTU table was rarefied by random sampling (without

replacement) at an even depth of 2000 reads to maximize the samples included in the analyses. On average, a total of 16,026 merged pair-end reads per sample were obtained after default quality filtering and chimera removal. The median amplicon length was 253bp after merging. These sequences resulted in a total of 7,137 OTUs. The major conclusions hold at different sequencing depths (Fig S1 in Supplemental information). Detailed information on the number of reads per sample is provided in Table S1 in Supplemental information.

The relative abundances of bacterial families and several measures of diversity were calculated using the rarefied OTU table. The number of OTUs, Shannon diversity index, and Faith's phylogenetic diversity (Faith 1992) were calculated. The number of OTUs provides a measure of species richness, while Shannon's diversity index provides an integrative measure of evenness and richness, and Faith's phylogenetic diversity is based on the cumulative phylogenetic branch lengths. For each sample, we calculated the mean of 20 iterations based on a subsampling of 2000 sequences. The differences in relative abundances of bacterial taxa and microbial diversity between the gut segments were tested using Kruskal-Wallis tests and Wilcoxon rank sum tests with Bonferroni corrections in JMP 12.1 (SAS institute). A two-tailed significance threshold of 0.05 was used for all the statistical tests.

To evaluate potential functional differences of the microbiome along the GI tract, gene family abundances were predicted from the 16S rRNA gene sequences using PICRUSt (Langille *et al.* 2013). OTUs that are only present in the reference database (Greengenes 13_5) were included in the analyses, a requirement of PICRUSt. The OTUs were rarefied to a single depth (2000 reads) for each sample. Kyoto Encyclopedia of Genes and Genomes (KEGG) ortholog abundances were predicted for each sample, the KEGG pathway functions were categorized at level 3, and the relative abundances of functional categories were calculated. Principle component values were calculated from all of the functional categories. Kruskal-Wallis tests and Wilcoxon rank sum tests with Bonferroni corrections were used to test for differences among gut segments.

Overall similarities between microbial communities were quantified using UniFrac distances which integrate the phylogenetic information of the community (Hamady, Lozupone and Knight 2010). Unweighted UniFrac distances (community membership; presence and absence of taxa) and weighted UniFrac distances (community structure; taking into account the relative abundances of taxa) were calculated in QIIME. To visualize the community similarity, PCoA and UPGMA hierarchical clustering analyses were conducted. Jackknife support of nodes in UPGMA trees were calculated using default settings in QIIME.

To quantify the effect size of variables explaining unweighted and weighted UniFrac distances, ADONIS was used with 999 permutations in QIIME. We also tested the effects of gut segment, individual, and diet (carbon and nitrogen stable isotopes) on gut microbial membership using Generalized Linear Models (GLMs) in R (version 3.3.0). To perform model selection, we used the Akaike information criterion with sample size correction (AICc) with the "AICc" function in the package "AICcmodavg", as well as likelihood ratio tests in cases where models were hierarchically nested using the "lrtest" function in the package "lmtest". We ran four separate analyses: (1) a full model, including fixed effects for gut segment, carbon, nitrogen, and individual; (2) fixed effects of carbon, nitrogen, and individual; (3) fixed effects of gut segment and individual; and

(4) fixed effects of gut segment, carbon, and nitrogen. Models 2, 3, and 4 are each nested within model 1 and were compared to model 1 in likelihood ratio tests. These tests evaluate whether gut segment, diet, and individual explain significant variation in microbial communities in the context of models that include the other variables. Each of these analyses was run on each of three datasets: the complete GI data, the upper GI data, and the lower GI data, for a total of twelve analyses (see Table S2 in Supplemental information). For each dataset, the response variable was unweighted UniFrac PC1, which was calculated independently in the complete GI, the upper GI, and the lower GI. The fractions of the variation explained by PC1 were 24%, 13%, and 22%, respectively.

3.3. Results

3.3.1. Obligately anaerobic bacteria are more abundant in the lower GI tract than they are in the upper GI tract

The relative abundances of the dominant phyla showed significant differences among gut segments (Kruskal-Wallis test, Firmicutes $P = 0.004$, Bacteroidetes $P < 0.0001$, Proteobacteria $P < 0.0001$, Tenericutes $P = 0.001$, and Cyanobacteria $P = 0.02$). Dominant bacterial families showed a similar pattern, in which 8 out of the 10 highly abundant bacterial families displayed significant differences in relative abundance among the gut segments after Bonferroni corrections (Fig 1B and Table S3 in Supplemental information). The observed differences in the relative abundances of bacterial taxa along the GI tract were consistent with a decrease in oxygen concentration from the mouth to the anus. For example, the upper GI tract (mouth to ileum) was dominated by mostly facultatively anaerobic bacterial families such as Pasteurellaceae, Mycoplasmataceae, and Lactobacillaceae (Table S4 in Supplemental information). In contrast, the lower GI tract (cecum to feces) was dominated by mostly obligately anaerobic bacterial families (Table S5 in Supplemental information). For example, within the phylum Firmicutes, facultatively anaerobic Firmicutes (Class: Bacilli) were more abundant in the upper GI tract than in the lower GI tract (Upper GI mean: 0.21, Lower GI mean: 0.01, Wilcoxon rank sum test: $P < 0.0001$). In contrast, obligately anaerobic Firmicutes (Class: Clostridia) showed the opposite pattern (Upper GI mean: 0.04, Lower GI mean: 0.37, Wilcoxon rank sum test: $P < 0.0001$).

Greater spatial heterogeneity in the relative abundances of bacterial taxa was observed within the upper GI tract compared to the lower GI tract (Table S4 and S5 in Supplemental information). For example, the relative abundances of Pasteurellaceae and Mycoplasmataceae showed significant differences among the gut segments within the upper GI tract (Kruskal-Wallis test, $P < 0.005$). Pasteurellaceae were more abundant in mouth, esophagus, and stomach samples (mean relative abundance = 0.57) compared to small intestine samples (mean relative abundance = 0.08) (Fig1B and Table S4 in Supplemental information). In contrast, Mycoplasmataceae were less abundant in mouth, esophagus, and stomach samples (mean relative abundance = 0.003) and were more abundant in small intestine samples (mean relative abundance = 0.52) (Fig1B and Table S4 in Supplemental information). However, within the lower GI tract, none of the 10

most abundant bacterial families showed significant differences in relative abundance between the gut segments (Table S5 in Supplemental information).

A total of 42 core OTUs (OTUs that are present in all individuals for each gut segment) was observed in the total dataset of 7,137 OTUs (Table S6 in Supplemental information). The number of core OTUs was greater in the lower GI tract (mean = 17.6) compared to the upper GI tract (mean = 4.8) (Wilcoxon rank sum test, $P = 0.01$). The distal cecum had the greatest number of core OTUs of all the gut segments. Fecal samples shared a greater fraction of core OTUs with the lower GI tract (93.3%) compared to the upper GI tract (6.7%) (Table S6 in Supplemental information).

3.3.2. Microbial diversity and community phylogenetic measurements differ between upper and lower GI tract

Microbial diversity differed among the gut segments overall (Kruskal-Wallis test, $P < 0.0001$ for Shannon index and Phylogenetic diversity). This result was mainly driven by the significant shifts in microbial diversity between the upper and lower GI tract. The Shannon index and Phylogenetic diversity were lower in the upper GI tract compared to the lower GI tract (Fig 1C) (Wilcoxon rank sum test: $P < 0.0001$ for each).

Although diversity measures were calculated after rarefying the number of sequences to an equal depth to control for sampling effort, both the Shannon index and Phylogenetic diversity were significantly correlated with the number of raw sequence reads in each sample (Fig S2 in Supplemental information). In principle, this might lead to a bias in estimates of diversity. However, three lines of evidence suggest that our conclusion of lower diversity in the upper GI tract is robust. First, when we restricted the comparison to samples where the read depths were comparable between upper and lower GI tract (i.e. samples with between 11634 and 18531 reads, Fig S2 in Supplemental information), the lower GI tract had consistently higher diversity measures compared to the upper GI tract (Wilcoxon rank sum test: $P < 0.0001$). Second, we took the residual values between the sequence reads and the rarefied diversity measures and we compared these residuals among gut segments; the overall pattern remained the same between the upper and lower GI tract using these residual values (Wilcoxon rank sum test: $P < 0.0001$) (Fig S3 in Supplemental information). Finally, comparisons at different sequence depths gave a consistent pattern using the rarefaction curves (Fig S1 in Supplemental information). Therefore, we conclude that microbial diversity is greater in the lower GI tract compared to the upper GI tract.

3.3.3. Predicted gene functions differ between the upper and lower GI tract

The predicted functions of the gut microbial community differed between the upper and lower GI tract. The first principle component of the relative abundances of microbial gene functions (KEGG pathway categories) predicted from the 16S rRNA gene data showed a significant difference between the upper and lower GI tract (Wilcoxon rank sum test $P < 0.0001$) (Fig 1D). The gene function PC1 was enriched in metabolism functions, where 11 out of the top 15 eigenvectors were categorized to functions related to metabolism (Table S7 in Supplemental information). Gene function PC1 significantly differed among the gut segments within the upper GI tract (Kruskal-Wallis test, P =

0.006), but not within the lower GI tract (Kruskal-Wallis test, P = 0.14). The differences among the gut segments in the upper GI tract were not significant after multiple corrections (Wilcoxon rank sum test, Bonferroni corrected P < 0.005). To test whether the abundance of genes in various metabolic pathways differed between the upper and lower GI tract, we focused on the top 15 most abundant metabolism gene function categories. Most of the metabolism gene function categories showed significant differences in their relative abundances between the upper and lower GI tract (Fig 2).

3.3.4. The effects of individual host, gut segment, and diet on microbial community composition

For the complete GI tract dataset, gut segment best explained the variation in microbial community membership ($R^2 = 0.31$, $P < 0.001$) (Fig 3A) and structure ($R^2 = 0.76$, $P < 0.001$) based on ADONIS (Table 1). Analyses using generalized linear models (GLMs) showed a similar pattern (Table S2 in Supplemental information). The model excluding gut segment (model 2) had a significantly lower log likelihood score compared to the full model (Likelihood ratio test, $P < 0.001$). In contrast, the model excluding diet (model 3) and the model excluding individual (model 4) were not significantly different from the full model (Table S2 in Supplemental information). This indicates that variation in diet and variation among individuals are not contributing to variation in gut microbial composition when the entire GI tract is considered. Given the large differences in microbial communities between the upper and lower GI tract, we next restricted our analyses to datasets consisting of samples from the upper GI tract alone and the lower GI tract alone.

Within the upper GI tract, gut segment was associated only with community structure (weighted UniFrac distance) ($R^2 = 0.66$, $P < 0.001$), but not with community membership (unweighted UniFrac distance) ($R^2 = 0.15$, $P = 0.2$) (Fig 3B and Table 1). Instead, community membership was significantly associated with host individual ($R^2 = 0.28$, $P < 0.001$) (Table 1). The model comparison of GLMs showed a similar pattern; a model without individual (model 4) was significantly worse than the full model (Likelihood ratio test, $p < 0.001$). In contrast, models excluding gut segment (AICc = -19.96) or excluding diet (AICc = -12.53) were comparable to the full model (AICc = -12.53) (Table S2 in Supplemental information). Within the lower GI tract, variation in community membership and variation in community structure were both significantly associated with host individual and diet, but not with gut segments (Fig 3C and Table 1). Similarly, a model without individual (model 4) was significantly worse than the full model (Likelihood ratio test, $p < 0.001$) (Table S2 in Supplemental information). Models without gut segment (AICc = -94.81) or without diet (AICc = -91.61) were comparable to or better than the full model (AICc = -19.61) (Table S2 in Supplemental information). These results indicate that differences among individuals account for significant variation in the gut microbial community when the upper or lower GI tract are considered separately. Moreover, these differences among individuals explain more of the variation in the lower GI tract than in the upper GI tract (Fig 3C, Table 1, and Table S2 in Supplemental information).

The stronger effect of individuals on microbial variation in the lower GI tract is further supported by UPGMA trees of community membership, where the samples

(cecum to feces) from the same individual were each clustered with a jackknife support of 1.0, unlike the pattern seen in samples from the upper GI tract (Fig 4). Although the lower GI tract samples from the same individual did not always form a clade when the taxa were weighted by relative abundances, the UPGMA tree based on community structure also showed a similar trend (Fig S4 in Supplemental information). Samples collected from the same geographic site grouped individuals in some cases (Female 1 and Male 1) but not in others (Female 2 and Male 2) (Fig 4). Although the sample size is very small, there seemed to be no obvious associations between geographic site and diet measures in the current dataset (Fig S5 in Supplemental information).

3.4. Discussion

We characterized the microbial communities from 10 locations along the GI tract in wild-caught house mice. We evaluated the relative importance of gut segment, individual host, and diet (as reflected in stable isotope measurements) in shaping differences in microbial composition. We found significant differences in microbial composition both among individuals and among gut segments. The effect of gut segments was most pronounced between the upper and lower GI tract. When these major divisions were considered separately, individual gut segments within each major division showed few differences, and differences among individual hosts showed the strongest effects on microbial composition. This suggests that individual-specific communities exist within both the upper and the lower GI tract. Below we discuss potential mechanisms underlying differences in microbial composition along the GI tract.

Observed differences in the relative abundance of anaerobic bacteria between the upper and lower GI tract were consistent with microbial oxygen requirements in humans (Hayashi *et al.* 2005) and in lab mice (Gu *et al.* 2013). Oxygen concentrations show a clinal decrease from the proximal to the distal GI tract in mice (He *et al.* 1999). The observed distribution of anaerobic microbial taxa can be explained by this oxygen gradient. Microbial oxygen requirements have been used as an explanation of microbial distribution in mucosal and luminal samples in mice, humans, and macaques (Espey 2013; Albenberg *et al.* 2014; Yasuda *et al.* 2015). However, we observed discrete shifts in the relative abundances of anaerobic taxa along the GI tract despite the gradual decrease in oxygen concentration.

Gut anatomy may help explain the discrete shifts in the distribution of anaerobic microbes along the GI tract. For example, house mice are hindgut fermenters, and most anaerobic fermentation takes place in the cecum, a pouch separating the small and large intestines. The cecum is where we observed an increase in anaerobic taxa. Others have made similar observations in hindgut fermenting rodents (Gu *et al.* 2013; Lu *et al.* 2014). In contrast, woodrats are foregut fermenters characterized by a segmented stomach in which fermentation takes place. In woodrats, anaerobic taxa were abundant throughout the GI tract including the stomach (Kohl *et al.* 2014a). Therefore, the distribution of anaerobic taxa may in part be determined by the particular anatomy associated with different kinds of fermentation chambers. It is also possible that anaerobic taxa were overrepresented in the lower GI tract compared to the upper GI tract due to differences in luminal *versus* mucosal biomass along the GI tract. Alternative explanations for the

distribution of microbial taxa are certainly possible (e.g. nutritional gradient, cell densities, immunity, etc.) and these explanations are not mutually exclusive. Characterizing the biochemical environment among the gut segments will aid in understanding the factors structuring the observed microbial communities.

Microbial diversity was greater in the lower GI tract compared to the upper GI tract. This pattern is consistent with studies in flying squirrels and lab mice (Gu *et al.* 2013; Lu *et al.* 2014) but different from humans and woodrats. In humans, mouth samples had the highest diversity (Stearns *et al.* 2011), and in woodrats, segmented stomach samples (where foregut fermentation takes place) were as diverse as the lower GI tract samples (Kohl *et al.* 2014a). The differences between humans and house mice could partly be explained by differences in sampling methods and by the greater sequencing depth in humans (Stearns *et al.* 2011). Although mouth and esophagus samples had smaller biomass compared to the rest of the GI tract in the current study, the initial biomass differences alone cannot explain the overall pattern because the rest of the gut segments were of equal weight. In rodents, the location of the fermentation chambers (the segmented stomach and/or cecum) along the GI tract might explain the diversity pattern; the small intestine shows the least diversity and the fermentation chambers show the highest diversity (Gu *et al.* 2013; Kohl *et al.* 2014a; Lu *et al.* 2014).

Natural selection may favor higher microbial diversity in fermentation chambers. Lu et al. (2014) suggested that the high diversity in fermentation chambers might reflect the stability-diversity relationship observed in macro-ecology (Tilman and Downing 1994; Doak *et al.* 1998): species-rich communities in fermentation chambers may be more stable, resilient, and recover faster from disturbance. Interestingly, herbivorous mammals also have the highest gut microbial diversity compared to omnivorous and carnivorous mammals (Ley *et al.* 2008b) suggesting the need for high microbial diversity for plant digestion. However, diversity measures should be interpreted with caution when different sample types are compared, since the sampled luminal volume or total cell numbers are likely to be different among the gut segments. Also our data cannot distinguish live cells from dead cells. Quantifying the number of cells sampled from each gut segment, and the viability of the cells will provide a better estimate of the microbial diversity across the GI tract.

Gut microbes in the lower GI tract showed a stronger pattern of individual-specific communities compared to the upper GI tract. The harsher environment in the upper GI tract (e.g. stomach acids) may filter certain bacterial taxa and potentially reduce the individual variation in the upper GI tract compared to the lower GI tract. Alternatively, individual differences in immunity in the lower GI tract might increase the individual variation in the lower GI tract compared to the upper GI tract (Mowat and Agace 2014). Understanding the effects of hosts on microbial composition is challenging for several reasons. The present study is limited not only due to the small sample size, but also due to the co-variation of multiple factors including diet, geographic site captured, sex, and host genotype. Manipulative experiments in a common environment would help characterize the effect of each variable in structuring the individual differences in microbial communities along the GI tract.

Microbiota of fecal samples were similar to those from the lower GI tract. Fecal samples had a comparable relative abundance of bacterial families, diversity measures, and predicted gene functions in comparison to those from the lower GI tract. Most of the

core OTUs (93.3%) found in the fecal samples were present in the lower GI tract samples. In terms of community membership and structure, fecal samples were indistinguishable from those derived from the lower GI tract. Surprisingly, the fecal samples accurately predicted the microbial community in the lower GI tract of each individual despite the fact that some individuals in this study were captured in the same geographic site and were eating a similar diet based on isotope diet measures. Although direct sampling from the gut segments of interest is ideal when possible, fecal samples are easy to collect, non-disruptive, comparable to previous studies, and are required for some longitudinal experiments. The similar community composition between the lower GI tract and feces, and the stronger effect of individuals in the lower GI tract compared to the upper GI tract support the utility of fecal sampling for studying gut microbial communities.

We characterized substantial heterogeneity among segments of the GI tract in wild house mice. However, individual hosts also play a significant role in structuring microbial communities within particular segments of the GI tract. Further research is required to understand the specific factors affecting the microbial community composition among gut segments and among individuals.

3.5. Chapter 3 Tables

Table 1. Variables explaining gut microbial communities based on ADONIS.

	Complete GI tract				Upper GI tract				Lower GI tract			
	Community membership[1]		Community structure[2]		Community membership[1]		Community structure[2]		Community membership[1]		Community structure[2]	
	R^2	p-values	R^2	p-values	R^2	p-values	R^2	p-values	R^2	p-values	R^2	p-values
Gut segments	**0.31**	**0.001**	**0.76**	**0.001**	0.15	0.166	**0.66**	**0.001**	0.10	0.995	0.10	0.807
Individual	**0.18**	**0.001**	0.09	0.441	**0.28**	**0.001**	0.14	0.658	**0.56**	**0.001**	**0.74**	**0.001**
Diet (δ^{13}C‰)	**0.04**	**0.020**	0.03	0.192	**0.06**	**0.003**	0.06	0.182	**0.09**	**0.007**	**0.23**	**0.001**
Diet (δ^{15}N‰)	**0.05**	**0.004**	0.03	0.140	**0.06**	**0.003**	0.03	0.347	**0.15**	**0.001**	**0.31**	**0.001**

[1] Unweighted UniFrac distance, which does not depend on relative abundance.
[2] Weighted UniFrac distance, which does depend on relative abundance.

3.6. Chapter 3 Figures

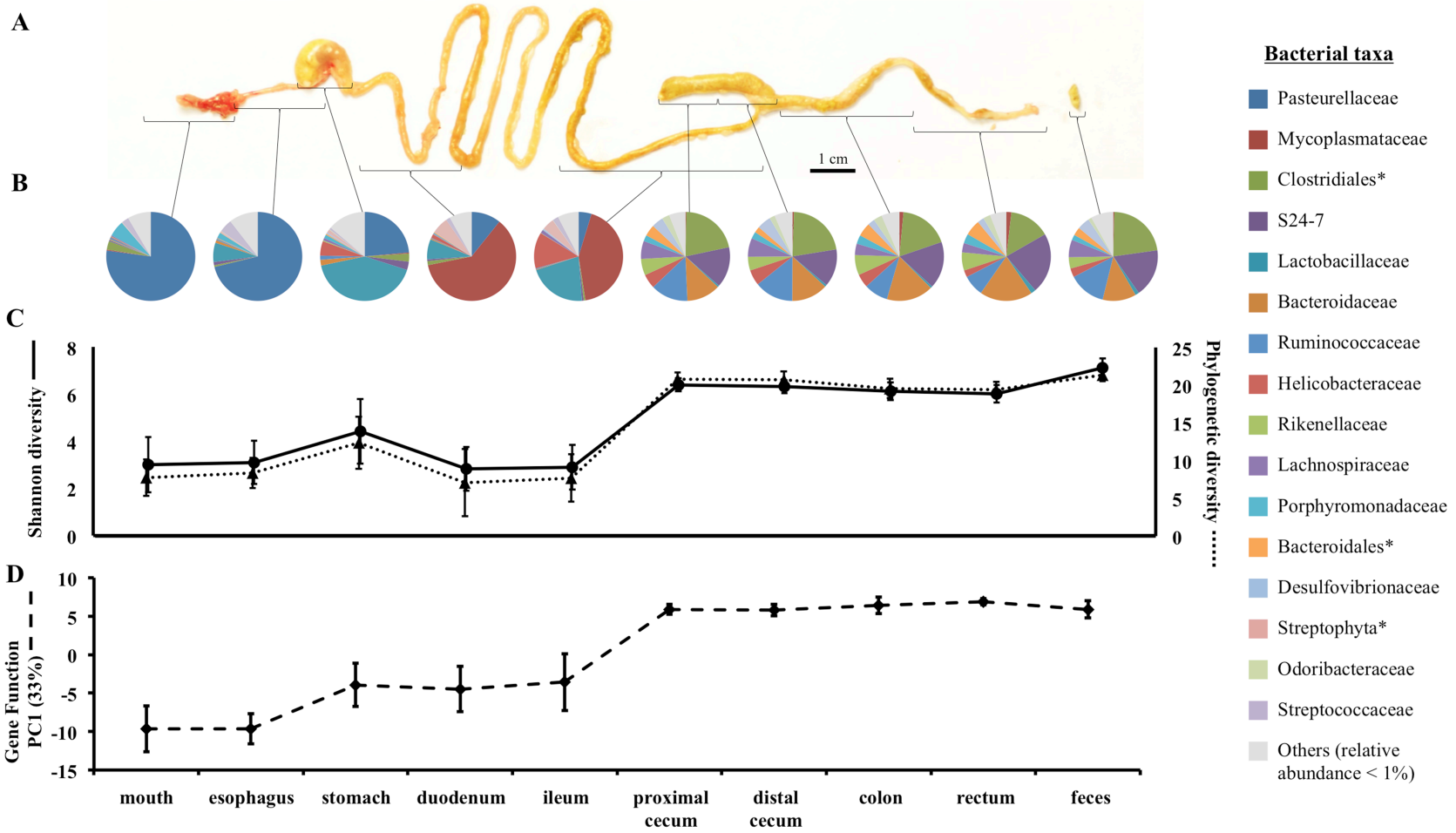


Fig 1. Spatial heterogeneity of microbial composition along the gastrointestinal (GI) tract. Ten samples were collected along the GI tract per individual (A). The averages of relative abundance of bacterial families (B), the microbial diversity measures (C), and the first principle component value from predicted gene function categories (D) across the GI tract are shown. The error bars are standard deviations. Asterisks denote unclassified families within the listed order.

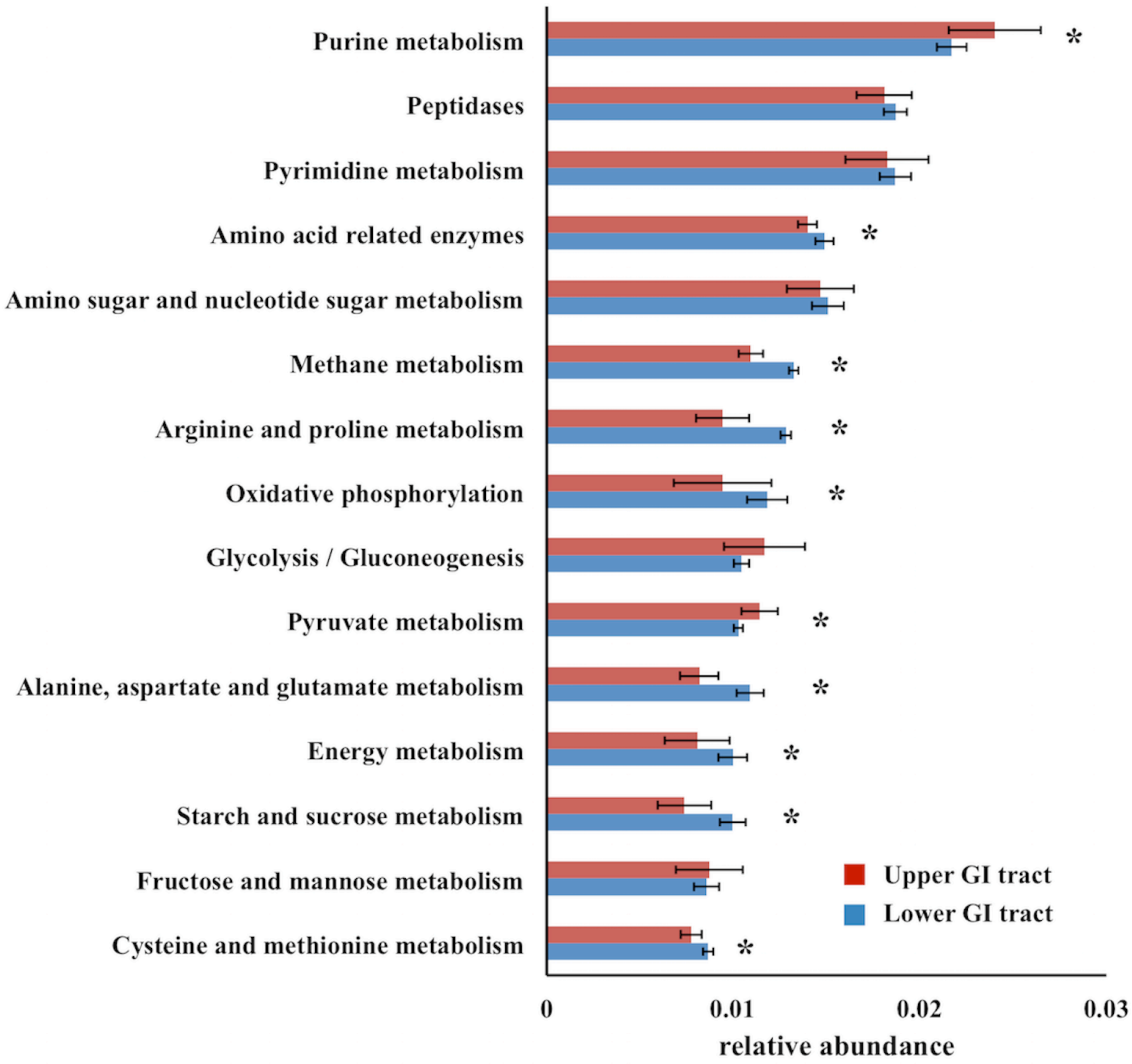


Fig 2. The relative proportions of the most abundant metabolism related KEGG pathways (level 3) predicted by PICRUSt between upper and lower GI tract. The error bars are standard deviations. The star indicates Bonferroni corrected $P < 0.0033$ using Wilcoxson rank sum test.

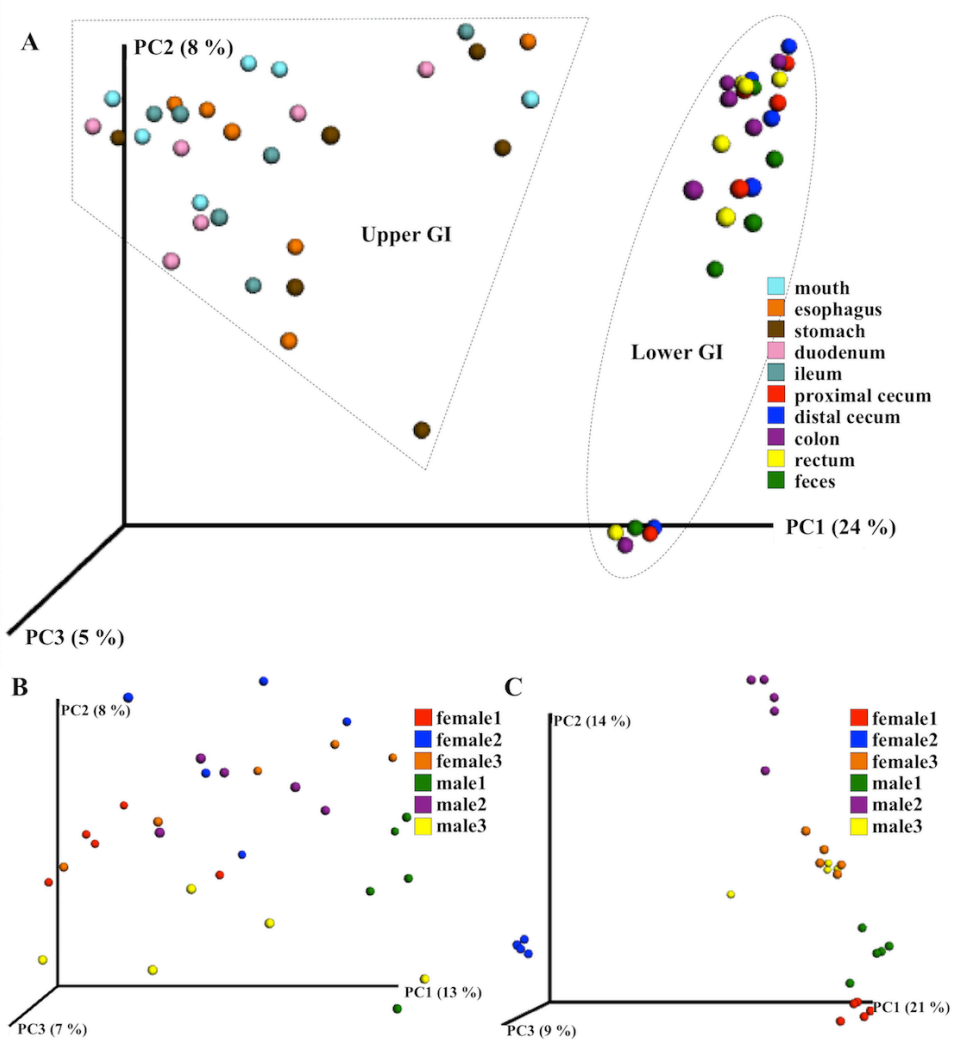


Fig 3. PCoA plots of microbial community membership (unweighted UniFrac distance). Each dot represents a microbial community from one gut segment in one individual. The first principle component (PC1) mostly accounts for differences between the upper and lower GI tract (A). Within the upper GI tract, microbial communities were grouped by host individual to some degree (B). Within the lower GI tract, microbial communities were strongly grouped by host individual (C).

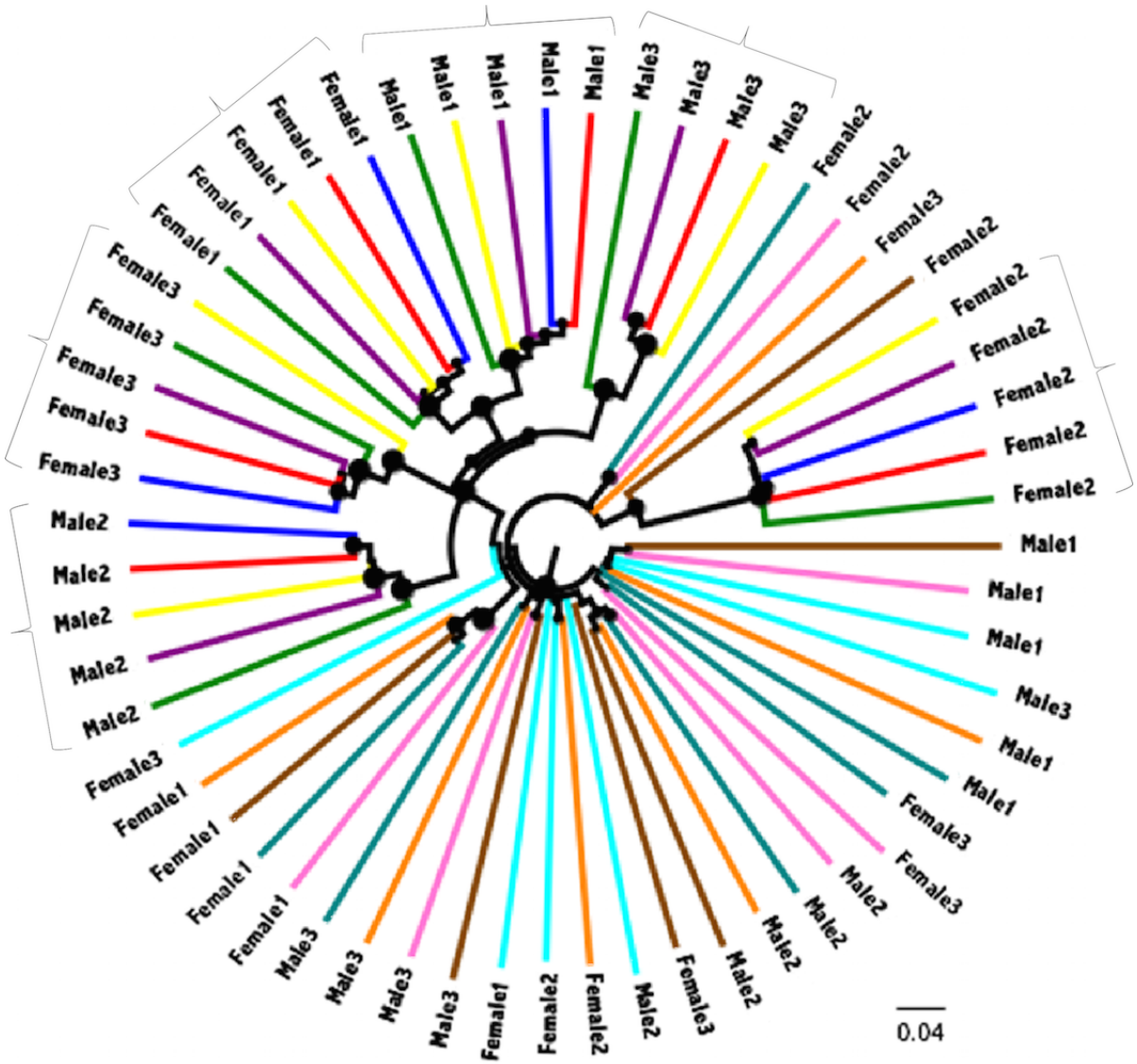


Fig 4. Microbial community membership is strongly associated with host individual in the lower GI tract. Tree is based on UPGMA clustering of unweighted UniFrac distance. Different colors show different gut segments (see Fig 3A). Larger node sizes indicate stronger jackknife support. The brackets show the clustering by individuals in the lower GI tract with a jackknife support of 1.0.

3. 7. Chapter 3 Supplemental information

Table S1. Sample information.

ID	GI location	Filtered reads[1]	Included in analyses?	Sex	ID in paper	δ13C ‰	δ15N ‰	Latitude	Longitude	Captured locality	Arctos ID	Sequence Sample IDs (ENA Accession PRJEB15238)
TAS.194.1	Mouth	15254	Yes	M	Male1	-18	8.7	32.174	111.01	Location 1 : Arizona, Pima Co., 2509 W. Zinnia Ave.	MVZ230525	TAS_ENA121
TAS.194.2	Esophagus	9943	Yes									TAS_ENA122
TAS.194.3	Stomach	12117	Yes									TAS_ENA123
TAS.194.4	Duodenum	17187	Yes									TAS_ENA124
TAS.194.5	Ileum	18531	Yes									TAS_ENA125
TAS.194.6	Distal Caecum	23251	Yes									TAS_ENA126
TAS.194.7	Proximal Caecum	20575	Yes									TAS_ENA127
TAS.194.8	Colon	21309	Yes									TAS_ENA128
TAS.194.9	Rectum	25201	Yes									TAS_ENA129
TAS.194.10	Feces	24592	Yes									TAS_ENA130
TAS.195.1	Mouth	15750	Yes	F	Female1	-20	10.3	32.174	111.01	Location 1 : Arizona, Pima Co., 2509 W. Zinnia Ave.	MVZ230526	TAS_ENA131
TAS.195.2	Esophagus	14925	Yes									TAS_ENA132
TAS.195.3	Stomach	15350	Yes									TAS_ENA133
TAS.195.4	Duodenum	14514	Yes									TAS_ENA134
TAS.195.5	Ileum	13032	Yes									TAS_ENA135
TAS.195.6	Distal Caecum	20365	Yes									TAS_ENA136
TAS.195.7	Proximal Caecum	24801	Yes									TAS_ENA137
TAS.195.8	Colon	22263	Yes									TAS_ENA138
TAS.195.9	Rectum	22952	Yes									TAS_ENA139
TAS.195.10	Feces	22436	Yes									TAS_ENA140
TAS.200.1	Mouth	5493	Yes	F	Female2	-20	6.2	32.1	111.05	Location 2: Arizona, Pima Co., 5665 W Ajo Hwy	MVZ230535	TAS_ENA141
TAS.200.2	Esophagus	5505	Yes									TAS_ENA142
TAS.200.3	Stomach	15909	Yes									TAS_ENA143
TAS.200.4	Duodenum	17873	Yes									TAS_ENA144
TAS.200.5	Ileum	16539	Yes									TAS_ENA145
TAS.200.6	Distal Caecum	21436	Yes									TAS_ENA146
TAS.200.7	Proximal Caecum	23087	Yes									TAS_ENA147
TAS.200.8	Colon	19749	Yes									TAS_ENA148
TAS.200.9	Rectum	20599	Yes									TAS_ENA149
TAS.200.10	Feces	20262	Yes									TAS_ENA150

[1] Reads after removing singletons and chimeras (reference based and de nov based chimeras).

Table S2. Sample information (continued).

ID	GI location	Filtered reads[1]	Included in analyses?	Sex	ID in paper	δ13C ‰	δ15N ‰	Latitude	Longitude	Captured locality	Arctos ID	Sequence Sample IDs (ENA Accession PRJEB15238)
TAS.201.1	Mouth	15334	Yes	M	Male2	-18	8.7	32.1	111.05	Location 2: Arizona, Pima Co., 5665 W Ajo Hwy	MVZ230537	TAS_ENA151
TAS.201.2	Esophagus	14525	Yes									TAS_ENA152
TAS.201.3	Stomach	4607	Yes									TAS_ENA153
TAS.201.4	Duodenum	9808	Yes									TAS_ENA154
TAS.201.5	Ileum	11065	Yes									TAS_ENA155
TAS.201.6	Distal Caecum	21695	Yes									TAS_ENA156
TAS.201.7	Proximal Caecum	18606	Yes									TAS_ENA157
TAS.201.8	Colon	20071	Yes									TAS_ENA158
TAS.201.9	Rectum	26219	Yes									TAS_ENA159
TAS.201.10	Feces	23159	Yes									TAS_ENA160
TAS.202.1	Mouth	13273	Yes	F	Female3	-20	8.5	32.08	111.04	Location 3: Arizona, Pima Co., 4517 W Valencia Rd.	MVZ230530	TAS_ENA161
TAS.202.2	Esophagus	12947	Yes									TAS_ENA162
TAS.202.3	Stomach	13585	Yes									TAS_ENA163
TAS.202.4	Duodenum	16484	Yes									TAS_ENA164
TAS.202.5	Ileum	2347	Yes									TAS_ENA165
TAS.202.6	Distal Caecum	11634	Yes									TAS_ENA166
TAS.202.7	Proximal Caecum	15059	Yes									TAS_ENA167
TAS.202.8	Colon	12193	Yes									TAS_ENA168
TAS.202.9	Rectum	14362	Yes									TAS_ENA169
TAS.202.10	Feces	17859	Yes									TAS_ENA170
TAS.203.1	Mouth	11713	Yes	M	Male4	-11	11.3	32.169	110.57	Location 4: Arizona, Pima Co., University of Arizona agricultural center	MVZ230539	TAS_ENA171
TAS.203.2	Esophagus	13280	Yes									TAS_ENA172
TAS.203.3	Stomach	11321	Yes									TAS_ENA173
TAS.203.4	Duodenum	6169	Yes									TAS_ENA174
TAS.203.5	Ileum	11224	Yes									TAS_ENA175
TAS.203.6	Distal Caecum	26153	Yes									TAS_ENA176
TAS.203.7	Proximal Caecum	215	No									TAS_ENA177
TAS.203.8	Colon	14177	Yes									TAS_ENA178
TAS.203.9	Rectum	15875	Yes									TAS_ENA179
TAS.203.10	Feces	15789	Yes									TAS_ENA180

[1] Reads after removing singletons and chimeras (reference based and de nov based chimeras).

Table S2. Model selection using Genelalized Linear Models on gut microbial membership (unweighted UniFrac PC1) along the GI tract.

Generalized Linear Models[1]		Complete GI tract		Upper GI tract		Lower GI tract	
	Fixed effects	AICc	Log likelihood[3]	AICc	Log likelihood[3]	AICc	Log likelihood[3]
Model 1	gut segment[2] + carbon + nitrogen + individual	-101.69	60.29	-12.53	24.60	-91.61	64.57
Model 2	carbon + nitrogen + individual	19.29	-1.55***	-19.96	19.52*	-94.81	57.07**
Model 3	gut segment[2] + individual	-101.69	60.29	-12.53	24.6	-91.61	64.57
Model 4	gut segment[2] + carbon + nitrogen	-104.65	57.89	-2.91	12.89***	-30.16	26.68***

[1] The response variable, unweighted UniFrac PC1 was calculated independently in Complete GI, Upper GI, and Lower GI tract. The fractions of the variation explained by PC1 are 24%, 13%, and 22%, respectively.

[2] Two parameters (upper GI and lower GI) were used for gut segment in "Complete GI tract" models. Five parameters (mouth, esophagus, stomach, dudenum, ileum) were used for gut segment in "Upper GI tract" models. Five parameters (proximal cecum, distal cecum, colon, rectum, feces) were used for gut segment in "Lower GI tract" models.

[3] Significant differences compared to the full model were determined based on the likelihood ratio test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table S3. Testing differences in relative abundance of dominant bacterial families between gut segments across the complete GI tract using Kruskal-Wallis test.

Complete GI tract taxa	Mean relative abundance	P-value[1]	Slope[2]
Proteobacteria; Pasteurellaceae	19.1%	**< 0.0001**	-
Tenericutes; Mycoplasmataceae	11.1%	*0.009*	-
Firmicutes; Clostridales*	10.6%	**< 0.0001**	+
Bacteroidetes; S24-7	9.0%	**< 0.0001**	+
Firmicutes; Lactobacillaceae	8.4%	**0.0007**	-
Bacteroidetes; Bacteroidaceae	7.7%	**< 0.0001**	+
Firmicutes; Ruminococcaceae	6.0%	**< 0.0001**	+
Proteobacteria; Helicobacteraceae	4.1%	*0.01*	+
Bacteroidetes; Rikenellaceae	3.0%	**< 0.0001**	+
Firmicutes; Lachnospiraceae	3.0%	**< 0.0001**	+

[1] Raw p-values are reported. P-values that remain significant after Bonferroni correction (alpha = 0.005) are bolded.

[2] Positive or negative slopes indicate decrease or increase in relative abundance of bacterial taxa from mouth to feces, respectively.

* unclassified family

Table S4. Testing differences in relative abundance of dominant bacterial families between gut segments within the upper GI tract using Kruskal-Wallis test.

Upper GI tract taxa	Mean relative abundance	P-value[1]	Slope[2]
Proteobacteria; Pasteurellaceae	37.5%	**0.0002**	-
Tenericutes; Mycoplasmataceae	20.9%	**0.0049**	+
Firmicutes; Lactobacillaceae	15.5%	*0.0064*	+
Proteobacteria; Helicobacteraceae	4.0%	0.09	NA
Cyanobacteria; Streptophyta*	2.4%	0.7	NA
Bacteroidetes; Streptococcaceae	2.3%	0.2	NA
Bacteroidetes; Porphyromonadaceae	1.9%	*0.02*	-
Firmicutes; Clostridales*	1.7%	0.6	NA
Bacteroidetes; Weeksellaceae	1.3%	*0.01*	-
Bacteroidetes; S24-7	1.2%	0.11	NA

[1] Raw p-values are repored. P-values that remian significant after Bonferroni correction (alpha = 0.005) are bolded.

[2] Positive or negative slopes indicate decrease or increase in relative abundance of bacterial taxa from mouth to ileum, respectively. Slopes for non-signifcant taxa were indicated by NA.

* unclassified family

Table S5. Testing differences in relative abundance of dominant bacterial families between gut segments within the lower GI tract using Kruskal-Wallis test.

Lower GI tract taxa	Mean relative abundance	P-value[1]	Slope[2]
Firmicutes; Clostridales*	19.7%	0.8	NA
Bacteroidetes; S24-7	17.1%	0.5	NA
Bacteroidetes; Bacteroidaceae	14.8%	0.4	NA
Firmicutes; Ruminococcaceae	11.5%	0.07	NA
Bacteroidetes; Rikenellaceae	5.8%	0.7	NA
Firmicutes; Lachnospiraceae	5.4%	0.09	NA
Proteobacteria; Helicobacteraceae	4.2%	0.8	NA
Bacteroidetes; Bacteroidales*	4.1%	0.6	NA
Proteobacteria; Desulfovibrionaceae	3.7%	0.7	NA
Bacteroidetes; Porphyromonadaceae	2.7%	0.8	NA

[1] Raw p-values are repored. P-values that remian significant after Bonferroni correction (alpha = 0.005) are bolded.

[2] Positive or negative slopes indicate decrease or increase in relative abundance of bacterial taxa from proximal cecum to feces, respectively. Slopes for non-signifcant taxa were indicated by NA.

* unclassified family

Table S6. A list of OTUs present in all individuals for each gut compartment (100% core OTUs).

Gut compartments	OTU IDs	Phyla	Order	Family	Genus	Species
Mouth	4298224	Firmicutes	Lactobacillales	Streptococcaceae	*Streptococcus*	-
Mouth	744251	Firmicutes	Lactobacillales	-	-	-
Mouth	70728	Proteobacteria	Pasteurellales	Pasteurellaceae	*Aggregatibacter*	*pneumotropica*
Mouth	4352772	Proteobacteria	Pasteurellales	Pasteurellaceae	*Aggregatibacter*	-
Mouth	9610	Proteobacteria	Pasteurellales	Pasteurellaceae	*Mannheimia*	-
Mouth	4295455	Proteobacteria	Pasteurellales	Pasteurellaceae	-	-
Mouth	82728	Proteobacteria	Pasteurellales	Pasteurellaceae	-	-
Esophagus	4298224	Firmicutes	Lactobacillales	Streptococcaceae	*Streptococcus*	-
Esophagus	70728	Proteobacteria	Pasteurellales	Pasteurellaceae	*Aggregatibacter*	*pneumotropica*
Esophagus	4352772	Proteobacteria	Pasteurellales	Pasteurellaceae	*Aggregatibacter*	-
Esophagus	4352772	Proteobacteria	Pasteurellales	Pasteurellaceae	*Aggregatibacter*	-
Esophagus	82728	Proteobacteria	Pasteurellales	Pasteurellaceae	-	-
Stomach	135956	Firmicutes	Lactobacillales	Lactobacillaceae	*Lactobacillus*	-
Stomach	70728	Proteobacteria	Pasteurellales	Pasteurellaceae	*Aggregatibacter*	*pneumotropica*
Stomach	82728	Proteobacteria	Pasteurellales	Pasteurellaceae	-	-
Duodenum	135956	Firmicutes	Lactobacillales	Lactobacillaceae	*Lactobacillus*	-
Duodenum	4298224	Firmicutes	Lactobacillales	Streptococcaceae	*Streptococcus*	-
Duodenum	303652	Firmicutes	Lactobacillales	Streptococcaceae	*Streptococcus*	-
Duodenum	70728	Proteobacteria	Pasteurellales	Pasteurellaceae	*Aggregatibacter*	*pneumotropica*
Duodenum	82728	Proteobacteria	Pasteurellales	Pasteurellaceae	-	-
Ileum	135956	Firmicutes	Lactobacillales	Lactobacillaceae	*Lactobacillus*	-
Ileum	4298224	Firmicutes	Lactobacillales	Streptococcaceae	*Streptococcus*	-
Ileum	303652	Firmicutes	Lactobacillales	Streptococcaceae	*Streptococcus*	-
Ileum	70728	Proteobacteria	Pasteurellales	Pasteurellaceae	*Aggregatibacter*	*pneumotropica*
Proximal Cecum	264325	Bacteroidetes	Bacteroidales	Rikenellaceae	-	-
Proximal Cecum	271418	Bacteroidetes	Bacteroidales	S24-7	-	-
Proximal Cecum	174573	Bacteroidetes	Bacteroidales	S24-7	-	-
Proximal Cecum	258849	Bacteroidetes	Bacteroidales	S24-7	-	-
Proximal Cecum	264734	Bacteroidetes	Bacteroidales	S24-7	-	-
Proximal Cecum	185587	Firmicutes	Clostridiales	Lachnospiraceae	-	-
Proximal Cecum	263337	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-

- unclassified

Table S6. Continued.

Gut compartments	OTU IDs	Phyla	Order	Family	Genus	Species
Proximal Cecum	4357315	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Proximal Cecum	259699	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Proximal Cecum	263546	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Proximal Cecum	230268	Firmicutes	Clostridiales	Ruminococcaceae	*Ruminococcus*	-
Proximal Cecum	189840	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Proximal Cecum	267689	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Proximal Cecum	179547	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Proximal Cecum	4410988	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Proximal Cecum	263977	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Proximal Cecum	3957916	Firmicutes	Clostridiales	-	-	-
Proximal Cecum	163997	Firmicutes	Clostridiales	-	-	-
Proximal Cecum	3919797	Firmicutes	Clostridiales	-	-	-
Proximal Cecum	1684221	Proteobacteria	Desulfovibrionales	Desulfovibrionaceae	*Desulfovibrio*	-
Distal Cecum	575041	Bacteroidetes	Bacteroidales	Rikenellaceae	-	-
Distal Cecum	264325	Bacteroidetes	Bacteroidales	Rikenellaceae	-	-
Distal Cecum	175646	Bacteroidetes	Bacteroidales	S24-7	-	-
Distal Cecum	271418	Bacteroidetes	Bacteroidales	S24-7	-	-
Distal Cecum	258849	Bacteroidetes	Bacteroidales	S24-7	-	-
Distal Cecum	262148	Bacteroidetes	Bacteroidales	S24-7	-	-
Distal Cecum	264734	Bacteroidetes	Bacteroidales	S24-7	-	-
Distal Cecum	135956	Firmicutes	Lactobacillales	Lactobacillaceae	*Lactobacillus*	-
Distal Cecum	176868	Firmicutes	Clostridiales	Dehalobacteriaceae	*Dehalobacterium*	-
Distal Cecum	173892	Firmicutes	Clostridiales	Lachnospiraceae	-	-
Distal Cecum	263138	Firmicutes	Clostridiales	Lachnospiraceae	-	-
Distal Cecum	321484	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Distal Cecum	263337	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Distal Cecum	259699	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Distal Cecum	263546	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Distal Cecum	337727	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Distal Cecum	162005	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-

- unclassified

Table S6. Continued.

Gut compartments	OTU IDs	Phyla	Order	Family	Genus	Species
Distal Cecum	189840	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Distal Cecum	267689	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Distal Cecum	179547	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Distal Cecum	4410988	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Distal Cecum	266411	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Distal Cecum	3957916	Firmicutes	Clostridiales	-	-	-
Distal Cecum	163997	Firmicutes	Clostridiales	-	-	-
Distal Cecum	274546	Firmicutes	Clostridiales	-	-	-
Distal Cecum	1684221	Proteobacteria	Desulfovibrionales	Desulfovibrionaceae	*Desulfovibrio*	-
Colon	264325	Bacteroidetes	Bacteroidales	Rikenellaceae	-	-
Colon	271418	Bacteroidetes	Bacteroidales	S24-7	-	-
Colon	174573	Bacteroidetes	Bacteroidales	S24-7	-	-
Colon	258849	Bacteroidetes	Bacteroidales	S24-7	-	-
Colon	262148	Bacteroidetes	Bacteroidales	S24-7	-	-
Colon	264734	Bacteroidetes	Bacteroidales	S24-7	-	-
Colon	135956	Firmicutes	Lactobacillales	Lactobacillaceae	*Lactobacillus*	-
Colon	263337	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Colon	162005	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Colon	267689	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Colon	179547	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Colon	4410988	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Colon	3957916	Firmicutes	Clostridiales	-	-	-
Colon	163997	Firmicutes	Clostridiales	-	-	-
Colon	1684221	Proteobacteria	Desulfovibrionales	Desulfovibrionaceae	*Desulfovibrio*	-
Rectum	4449518	Bacteroidetes	Bacteroidales	Rikenellaceae	-	-
Rectum	264325	Bacteroidetes	Bacteroidales	Rikenellaceae	-	-
Rectum	175646	Bacteroidetes	Bacteroidales	S24-7	-	-
Rectum	271418	Bacteroidetes	Bacteroidales	S24-7	-	-
Rectum	174573	Bacteroidetes	Bacteroidales	S24-7	-	-
Rectum	258849	Bacteroidetes	Bacteroidales	S24-7	-	-

- unclassified

Table S6. Continued.

Gut compartments	OTU IDs	Phyla	Order	Family	Genus	Species
Rectum	262148	Bacteroidetes	Bacteroidales	S24-7	-	-
Rectum	264734	Bacteroidetes	Bacteroidales	S24-7	-	-
Rectum	135956	Firmicutes	Lactobacillales	Lactobacillaceae	*Lactobacillus*	-
Rectum	259699	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Rectum	267689	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Rectum	179547	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Feces	328617	Bacteroidetes	Bacteroidales	Bacteroidaceae	*Bacteroides*	*uniformis*
Feces	271418	Bacteroidetes	Bacteroidales	S24-7	-	-
Feces	174573	Bacteroidetes	Bacteroidales	S24-7	-	-
Feces	258849	Bacteroidetes	Bacteroidales	S24-7	-	-
Feces	262148	Bacteroidetes	Bacteroidales	S24-7	-	-
Feces	264734	Bacteroidetes	Bacteroidales	S24-7	-	-
Feces	135956	Firmicutes	Lactobacillales	Lactobacillaceae	*Lactobacillus*	-
Feces	263337	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Feces	259699	Firmicutes	Clostridiales	Ruminococcaceae	*Oscillospira*	-
Feces	267689	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Feces	179547	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Feces	4410988	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Feces	263977	Firmicutes	Clostridiales	Ruminococcaceae	-	-
Feces	3957916	Firmicutes	Clostridiales	-	-	-
Feces	1684221	Proteobacteria	Desulfovibrionales	Desulfovibrionaceae	*Desulfovibrio*	-

- unclassified

Table S7. Top 15 eigenvectors (PC loadings) for gene function PC1.

KEGG categories (level 1)	KEGG categories (level 3)	Eigenvectors for PC1 (33.2%)
Genetic information processing	Sulfur relay system	-0.14139
Metabolism	Glutathione metabolism	-0.13928
Metabolism	Phenylpropanoid biosynthesis	0.13874
Genetic information processing	Transcription machinery	0.13703
Metabolism	Metabolism of cofactors and vitamins	-0.1371
Metabolism	Cyanoamino acid metabolism	0.13582
Metabolism	Sphingolipid metabolism	0.13488
Metabolism	Alanine, aspartate, and glutamate metabolism	0.13469
NA*	Function unknown	-0.13458
Metabolism	Methane metabolism	0.13252
Metabolism	Other glycan degradation	0.13239
Metabolism	Glycosphingolipid biosynthesis - globo series	0.13175
Metabolism	Starch and sucrose metabolism	0.12869
Metabolism	One carbon pool by folate	0.12598
Human Diseases	Bacterial toxins	0.12531

* NA indicates not apply.

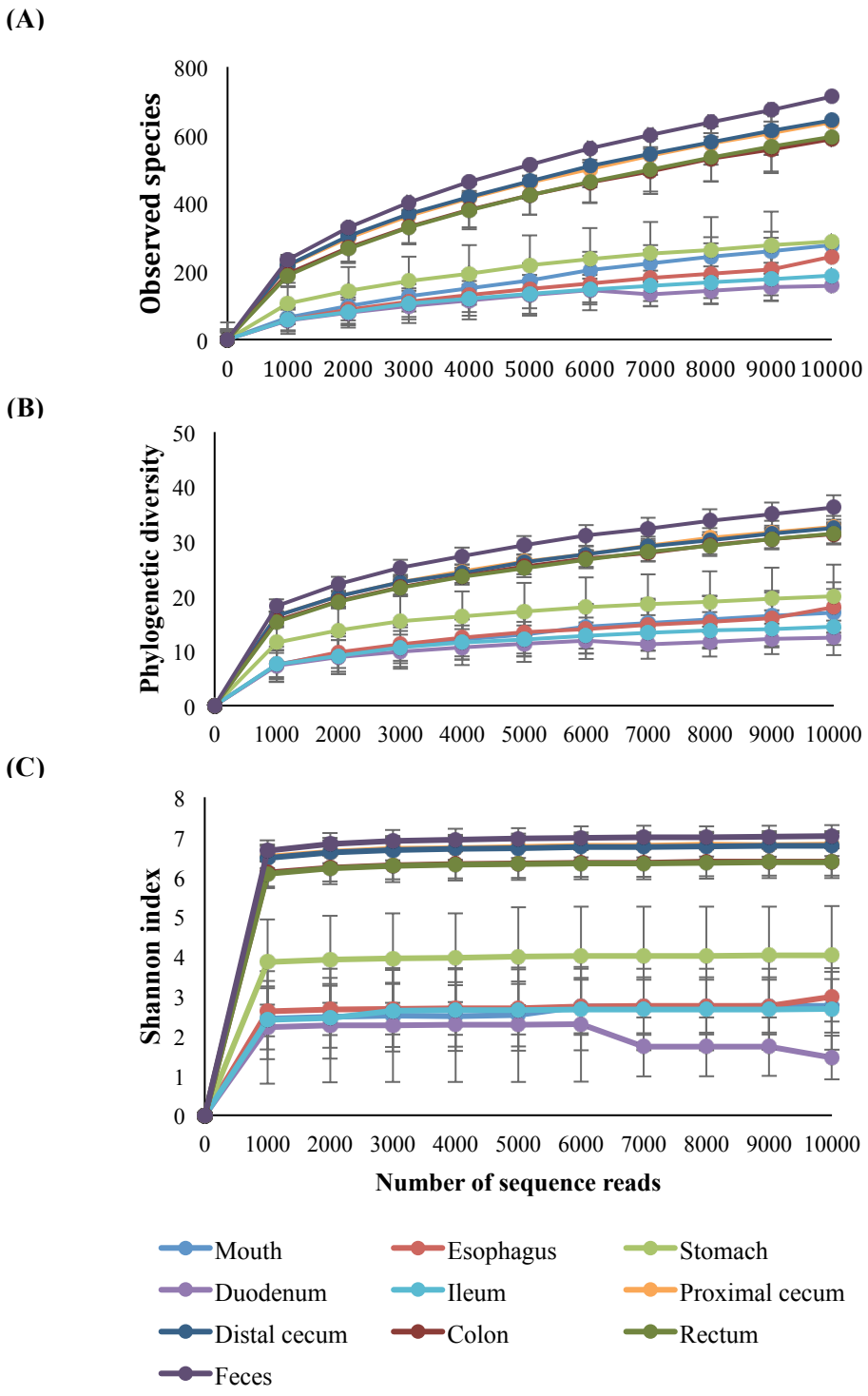


Figure S1. Rarefaction curves for Observed species (A), Phylogenetic diversity (B), and Shannon index (C). The error bars are standard deviations.

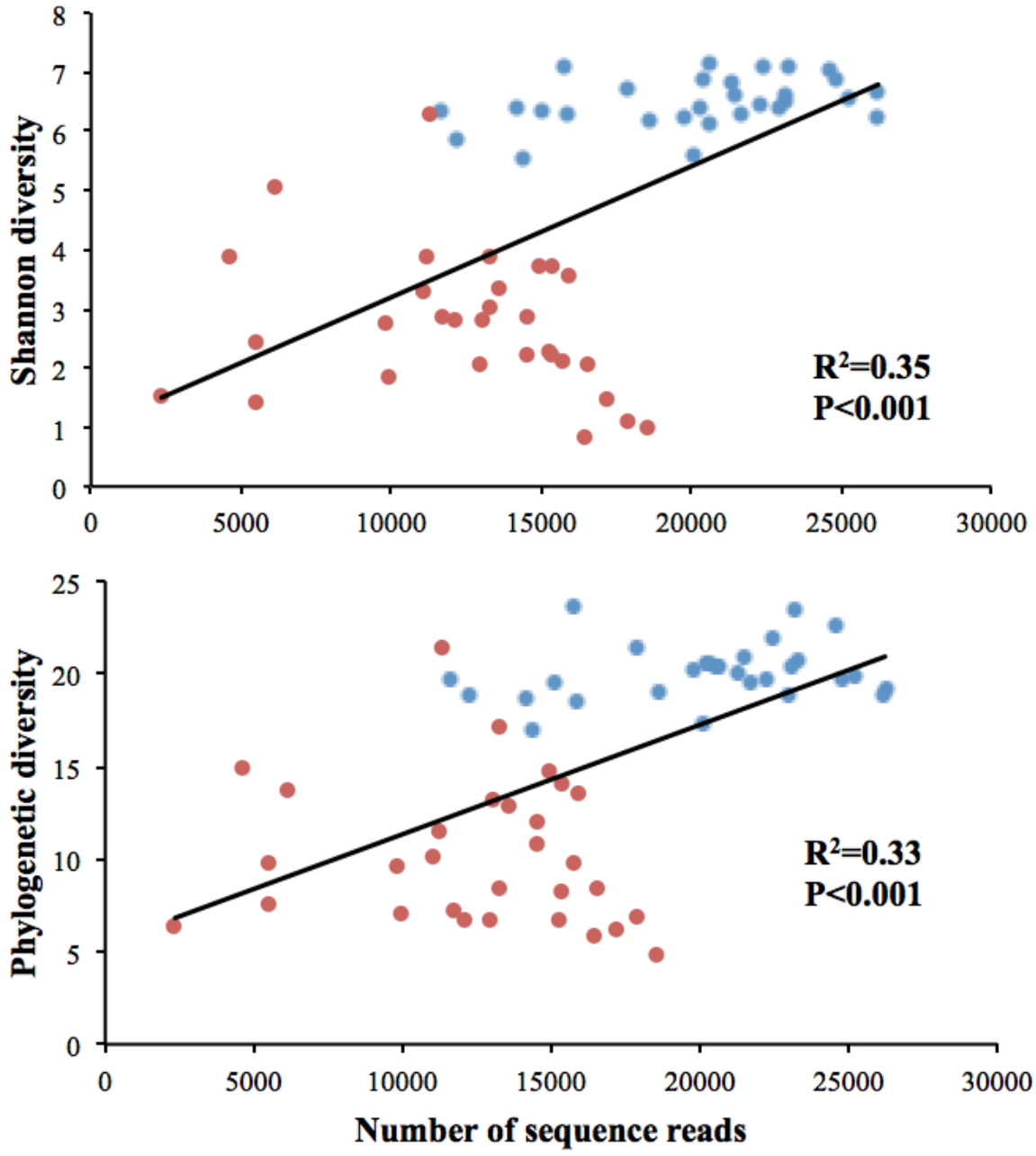


Figure S2. Linear regression between raw sequence reads and rarefied diversity measurements. Upper GI tract samples are shown in blue and lower GI tract samples are shown in red.

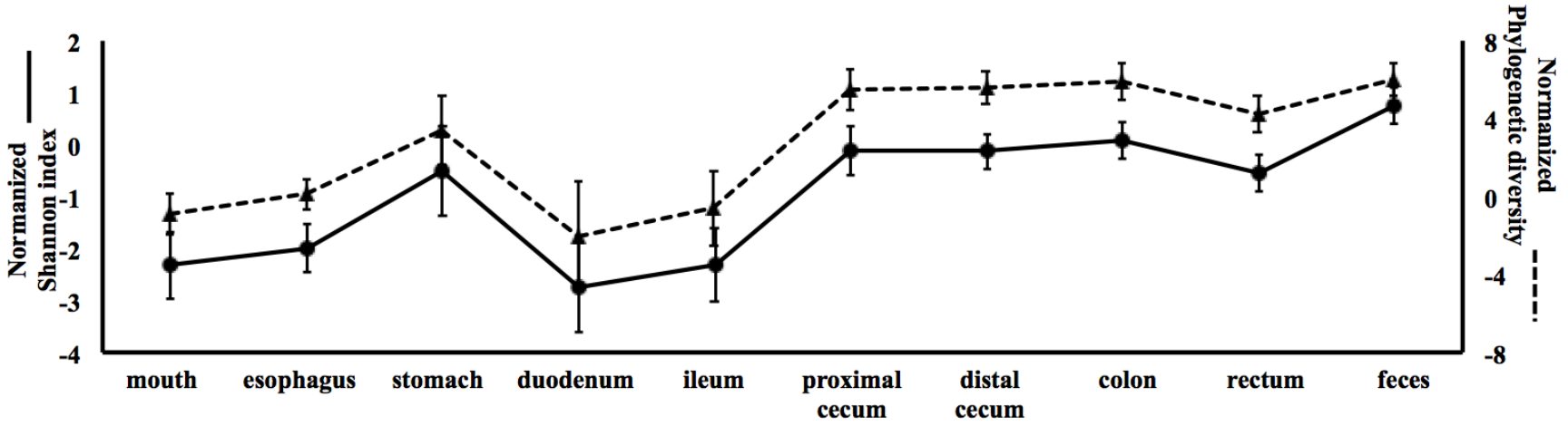


Figure S3. Normalized diversity measurements (residual values between raw sequence reads and rarefied diversity measurements) and gut segments. The normalized diversity measurements in lower GI tract are greater compared to upper GI tract (Wilcoxon rank sum test: $P<0.0001$), consistent with Fig.1C. Error bars are standard errors.

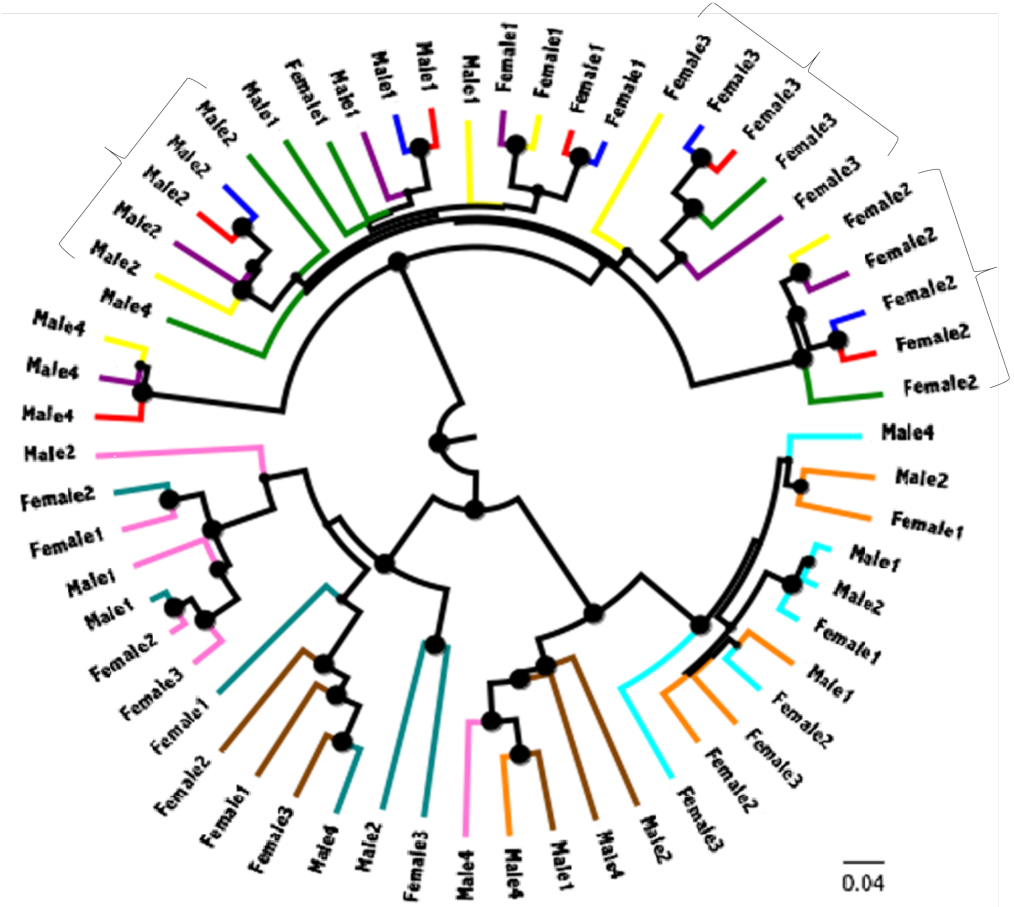


Figure S4. Stronger effect of individual explaining microbial structure membership in lower GI tract compared to upper GI tract based on UPGMA clustering. Different colors show different gut segments (see Fig.2). Larger node size indicates stronger jackknife support. The brackets show the clustering by individuals in lower GI tract.

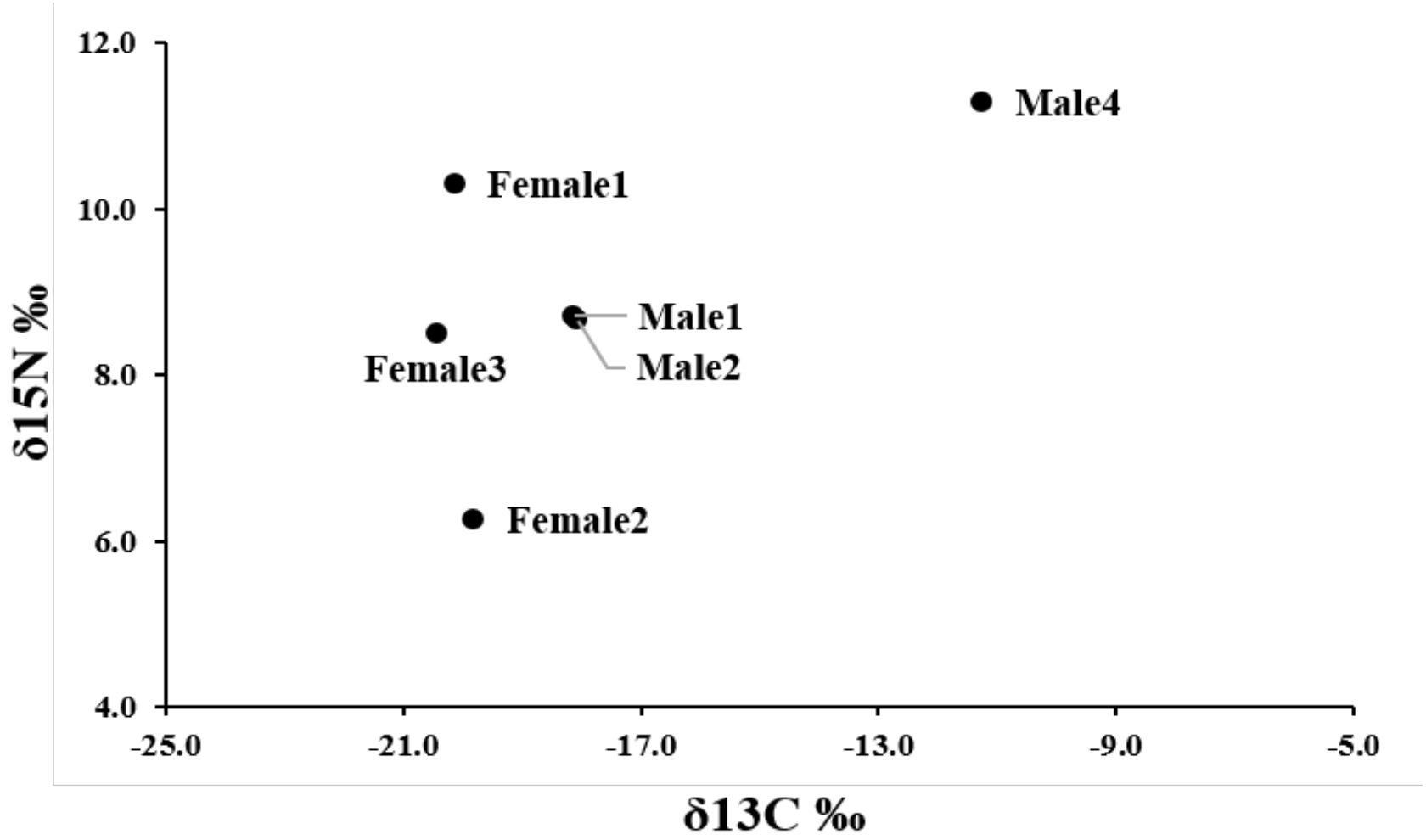


Figure S5. Carbon and nitrogen stable isotope diet measures from hair. No clear association between diet and geographic site captured. See Table S1 for captured locality.

Chapter 4

Host genetic determinants of the gut microbiota of wild mice

Anticipated co-authorship: Megan Phifer-Rixey[1], Katya Mack[2], Michael Sheehan[3], Ting-Ting Lin[2], Ke Bi[4], and Michael W. Nachman[2]

[1].Department of Biology, Monmouth University, West Long Branch, NJ, 07764, USA.
[2] Department of Integrative Biology and Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, CA, 94720, USA.
[3] Department of Neurobiology and Behavior, Cornell University, Ithica, NY, 14853, USA.
[4] California Institute for Quantitative Biosciences, University of California Berkeley, Berkeley, CA, 94720, USA.

Abstract

Identifying a common set of genes that mediate host-microbial interactions across populations and species of mammals has broad relevance for human health and animal biology. However, the genetic basis of the gut microbial composition in natural populations remains largely unknown outside humans. Here, we characterized variation in the gut microbiota of wild house mice using 16S amplicon sequencing and found that host genetic distance is a strong predictor of gut microbial composition. Using a common garden approach, we then identified differences in microbial composition between populations that persist in a shared laboratory environment. Finally, we used exome-sequencing to associate host genetic variants with microbial measurements in wild individuals. We identified 20 genes that were associated with bacterial measurements including a macrophage-derived cytokine (*IL12a*) that contained three nonsynonymous mutations. Surprisingly, we found a significant overrepresentation of candidate genes that were previously associated with microbial measurements in humans. All eight genes that overlapped between wild mice and humans were expressed in the brain, and some of the loci have also been associated with traits related to host immunity and obesity in humans. Gene-bacteria associations identified in both humans and wild mice suggest some commonality to the host genetic determinants of gut microbial composition across mammals.

4.1. Introduction

Host-associated microbial communities play an important role in health and fitness (McFall-Ngai *et al.* 2013). Compositional and functional variation in the gut microbiota has been linked to a variety of diseases in humans and lab mouse models including obesity, inflammatory bowel disease, and autism (Knight *et al.* 2017). Links between the gut microbiota and fitness-related traits have also been reported in wild mammals

including traits related to digestion, immunity, and behavior (Suzuki 2017). Therefore, understanding the mechanisms governing the maintenance and function of gut microbial communities is important in medicine and animal biology more broadly.

Genetic differences among hosts may play an important role in structuring gut microbial communities. For example, genome-wide markers have been associated with differences in microbial measurements in humans (Ma *et al.* 2014; Blekhman *et al.* 2015). Twin studies have shown that monozygotic twins tend to have more similar microbial composition compared to dizygotic twins (Goodrich *et al.* 2014, 2016a). Mouse knockout experiments have identified genes involved in immunity, metabolism, and behavior that affect the gut microbiota (Spor, Koren and Ley 2011). Mouse quantitative trait loci (QTL) mapping studies have also identified multiple genomic regions associated with the relative abundance of different microbial taxa (Benson *et al.* 2010; McKnite *et al.* 2012; Leamy *et al.* 2014; Org *et al.* 2015; Wang *et al.* 2015).

In human populations, microbiome genome-wide association studies (mGWAS) have identified specific candidate genes associated with natural variation of the gut microbiota (Knights *et al.* 2014; Blekhman *et al.* 2015; Davenport *et al.* 2015; Bonder *et al.* 2016; Goodrich *et al.* 2016a; Turpin *et al.* 2016; Wang *et al.* 2016). Although identifying the same gene-bacteria associations in different human populations has been challenging due to the low reproducibility of mGWAS, a few gene-bacteria associations have been replicated in multiple human populations (Goodrich *et al.* 2016b; Hall, Tolonen and Xavier 2017). This approach of comparing multiple mGWAS has been successful for identifying robust gene-bacteria associations within species, however, this approach has not been applied to look for gene-bacteria associations between species of mammals. Genes identified from human mGWAS are often compared with those identified in laboratory mice, but gene-bacteria associations identified in a controlled laboratory environment may differ from those in a complex natural environment. In fact, the function and composition of the gut microbiota in lab mice are known to differ from their wild relatives (Rosshart *et al.* 2017). Population samples from a wild species would provide an opportunity for mGWAS that are more directly comparable to human mGWAS.

Wild house mice (*Mus musculus domesticus*) are globally distributed and live in a wide range of environments in association with humans (Phifer-Rixey and Nachman 2015). The house mouse is a powerful model because it is possible to disentangle variables via experimental manipulation directly in captivity (Wang *et al.* 2014, 2015) and assess the functions of the microbiome by using germ-free laboratory mice (Rosshart *et al.* 2017). Previous work has shown that geographic and genetic distances (Linnenbrink *et al.* 2013), diet as measured by stable isotopes (Wang *et al.* 2014), reproductive status, body size, age, viral and parasite infection status (Weldon *et al.* 2015), and gut regions (Suzuki and Nachman 2016) have all been associated with the compositional differences in the gut microbiota of wild house mice. However, there have been no previous efforts to identify specific genes underlying compositional variation in the gut microbiome of wild mice.

Here, we characterized natural variation of the gut microbiota in wild house mice sampled from five populations along an environmental gradient in eastern North America using 16S amplicon sequencing, and we identified specific genes in mice associated with gut microbial composition. First, we tested how environmental factors and genetic

distance correlate with beta-diversity of the gut microbiota within and between populations. Second, we tested whether population differences in the wild mouse gut microbiota are due to environmental differences by conducting a common garden experiment in the laboratory. Third, we identified gene-bacteria associations in wild mice using a mGWAS with complete exome sequences of all mice. Finally, we tested whether there was significant overlap between the genes associated with microbiota variation in mice and genes associated with microbiota variation in humans.

4.2. Materials and Methods

4.2.1. Field collected animals

A total of 50 adult house mice (*Mus musculus domesticus*) were collected from the eastern North America during summer 2012 by Phifer-Rixey et al (*in review*) and summarized in Table S9. 10 individuals each were collected from a total of five populations; Florida (FL), Georgia (GA), Virginia (VA), Pennsylvania (PA), and spanning New Hampshire and Vermont (NH-VT). Sherman live traps were used with peanut butter and oats. Each individual was separated by a minimum 500m to avoid sampling close relatives. Animals were kept in Sherman traps, euthanized by cervical dislocation, and all tissues (liver, spleen, kidney, and cecum) and external measurements (body weight and body length) were collected within 24 hours after capture. All tissues were stored in liquid nitrogen in the field, and stored in deep freezer (-80°C) until sequencing. To infer diet, carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) stable isotope analyses were conducted using mouse hair following the protocol of Suzuki and Nachman (2016). To infer climate for each collection site, 19 climatic variables were downloaded based on GPS localities using WorldClim database (Hijmans *et al.* 2005) and the R package "dismo". First two principle components were calculated which accounts for 91.3% of the total variation (Table S10) using JMP 13.0 (SAS institute). All procedures involving animals were reviewed and approved by the Institutional Animal Care and Use Committee at the University of Arizona (protocol 07-004). Museum specimens (skins and skulls) were prepared and have been deposited in the mammal collection of the Museum of Vertebrate Zoology at the University of California, Berkeley and uploaded to a public database ARCTOS. See Phifer-Rixey et al. (*in review*) for more details.

4.2.2. Lab reared animals

Live animals were captured close to the most northern and southern populations using Sherman live traps during summer 2013; Saratoga Springs, NY and Gainsville, FL. Within each location, animals were collected from at least 10 sites that are minimum 500m apart. Animals were shipped to the animal facility at University of California Berkeley under IACUC protocol (R361-0514). Animals were kept in 23°C with 10 hours dark and 14 hour light cycles. Teklad Global food (18% Protein Rodent Diet) was fed *ad libitum*. After quarantine, we conducted 10 independent crosses per population where wild-caught individuals from the same site were not used in different crosses. Multiple individuals were housed in a cage (i.e. 1-3 individuals) with their siblings throughout

their life. Body weight and fresh fecal samples were collected from 40 individuals per population representing four adult individuals (two females and two males) each from 10 independent crosses of wild-caught founders. The age of the animals when fecal samples were collected ranged from 100 to 266 days old. Fecal samples were stored in -80°C until sequencing. Detailed information of lab-reared animals is in Table S11.

4.2.3. DNA extraction and 16S rRNA gene sequencing

For wild caught individuals, the distal portion of frozen cecal samples was weighted equally (~200mg) using a sterile razor blade in a petri dish on dry ice. For lab reared individuals, frozen fecal samples were weighted equally (~200mg). The samples were immediately mixed with sterile forceps in 1.4ml ASL (from QIAamp DNA stool Minikit). For mechanical disruption, 0.2g of sterile zirconia/silica beads (0.1mm, Research Products International Corp.) were added to the tubes and vortexed in TissueLayser LT (Qiagen) at 30HZ for 6 min (Smith *et al.* 2011). The suspension was heated at 95°C for 5 min, and we followed step 4 (vortex and centrifugation) in the protocol from the QIAamp DNA stool Minikit (Qiagen) to complete the DNA extraction. The DNA quality was evaluated based on Nano Drop 3300 (Thermo Scientific), and the samples were stored at -20°C before sequencing.

The extracted DNA samples were shipped to Next Generation Sequencing Core Facility at Argonne National Laboratory for 16S amplicon sequencing. The V4 region of the 16S rRNA gene was amplified and the samples were multiplexed. The PCR primers (515F and 806R) and the barcodes are described in (Caporaso *et al.* 2012). The samples from wild-caught mice and lab-reared mice ran on two different lanes of 150bp pair-end Illumina MiSeq platform. To avoid the potential lane bias between the two lanes, the same DNA aliquot of six samples from the first lane (MPR108, MPR114, MPR120, MPR135, MPR138, and MPR144) was run on the second lane.

4.2.4. 16S data processing

All of the 16S data were analyzed in QIIME version 1.9.0 (Caporaso *et al.* 2010). The forward reads were demultiplexed and quality-filtered using default parameters for each lane using *split_libraries_fastq.py*. Subsampled open-reference OTU picking approach (*pick_open_reference_otus.py*) was employed on two lanes of sequence data with default parameters. 97% OTUs were generated using UCLUST (Edgar 2010) and taxa were assigned based on Greengene database 13.8 (DeSantis *et al.* 2006). To remove sequence errors and very rare OTUs, OTUs with <10 reads across all samples were removed. A phylogenetic tree was created using FastTree (Price, Dehal and Arkin 2009). The OTU table was rarefied to even depth of 5,000 reads. Two samples (FL08M1 and FL08M2) were removed from all analyses due to low sequence reads (<200 reads).

Despite rarefying the reads to equal depth for all samples, the OTU counts were consistently higher in run2 compared to run1 for the six control samples (Fig. S6). This lane bias is likely due to the greater average sequence depth of run2 (68,196 reads per sample) compared to run1 (13,918 reads per sample) resulting in excess of rare OTUs in run2. To account for this, we removed rare OTUs from run2 to a degree that the control samples having equal OTU counts between the lanes (i.e. OTUs with relative abundance

less than $8.0x10^{-6}$ were removed) (Fig. S6). The OTU table corrected for the lane bias was used for all the downstream analyses.

4.2.5. Mouse exome data

We used the same exome data for the same individuals as in Phifer-Rixey et al. (*in review*). Briefly, the DNA was extracted from frozen liver, kidney, or spleen. Genomic libraries were enriched for mouse exome using a NimbleGen in-solution capture array (SeqCap EZ) and sequenced using 100bp pair-end Illumina HiSeq2000. After quality filtering and SNP discovery using custom perl script and PERL program, we filtered the SNPs based on minor allele frequency to be at least 5%. This resulted in 279,278 SNPs. Each SNP were annotated to a single or multiple genes using Variant Effect Predictor in Ensembl. See details of quality filtering and SNP discovery in Phifer-Rixey et al. (*in review*).

4.2.6. 16S data analyses

For beta-diversity measurements, Bray-Curtis dissimilarity, unweighted- and weighted-UniFrac distances were calculated among all individuals using *beta_diversity.py*. Distance matrix of predictor variables were created using *distance_matrix_from_mapping.py* for the following variables; geographic distance, body weight (g), body mass index (BMI), diet ($\delta^{13}C$), diet ($\delta^{15}N$), climate PC1, and climate PC2. Pairwise distances for geography (km) were calculated based on GPS coordinates of the captured site. Pairwise distances for host genomes were calculated based on the exome data (~280,000 SNPs) using ngsDist (Vieira *et al.* 2016) which takes into account uncertainty of the genotype calls.

To test correlations between beta-diversity and eight predictor variables across populations, we used Mantel test and Generalized Linear Models (GLMs) in R (version 3.3.2). To identify variables that significantly explain the major PC axes (PC1-3) of Bray-Curtis dissimilarity while controlling for all variables, we performed model selection using the Akaike information criterion with sample size correction (AICc) with the "AICc" function in the package "AICcmodavg" and likelihood ratio tests using the "lrtest" function in the package "lmtest". We ran nine separate GLMs each for Bray-Curtis PC1, 2, and 3 as response variables where models were hierarchically nested. A full model included fixed effects for genetic distance, geographic distance, body weight, BMI, carbon, nitrogen, climate PC1, and climate PC2. Eight other models included fixed effects where each variable was subtracted from the full model in turn. All eight models are each nested within the full model and were compared to the full model using likelihood ratio tests (Table S1). To test whether host genetic distance and geographic distance are independently associated with Bray-Curtis dissimilarity across populations while controlling the effects of each other, a partial mantel test was used. To ask the same question between individuals within populations, we used Spearman's rho correlation with residuals between Bray-Curtis dissimilarity and geographic distance to control for geography and residuals between Bury-Curtis dissimilarity and genetic distance to control for genetics. Similarly, to test correlations between beta-diversity and body size

measurements, Spearman's rho correlation with residuals between body size and latitude were used to control for the known effect of latitude on body size (Table S2).

To identify bacterial measurements that are associated with latitude, body weight, and BMI, we calculated alpha-diversity using *alpha_diversity.py* and relative abundances of bacterial taxa using *summarize_taxa.py* using the rarefied OTU table. For bacterial measurements, we used phylogenetic diversity (Faith 1992) as an alpha-diversity measurement and relative abundances of 17 bacterial genera that had an average relative abundance of >1% across all individuals which includes four Phyla. First, we tested the correlations between the bacterial measurements and latitude. Second, we tested the correlation between the microbial measurements and the residual values between body size measurements (i.e. body weight and BMI) and latitude to control for bacteria that is covaried with latitude (Fig. S4). Spearman's correlation with Bonferroni correction was used.

4.2.7. Common garden experiment

To confirm whether the observed host genetic associations with microbiota measurements are not due to environmental differences, we conducted a common garden experiment in the laboratory using the first-generation animals of wild-caught founders from the most northern and southern populations described above (Table S10). Although cecal and fecal samples were used to characterize the gut microbiota in wild and lab mice, respectively, the fecal samples have been shown to reflect individual differences in cecal samples in wild house mice (Suzuki and Nachman 2016).

We calculated Bray-Curtis dissimilarity among the northern-wild population (NH-VT), the southern-wild population (FL), the northern-lab population (NY), and the southern-lab population (FL). We tested the significance using Wilcoxon permutation test based on 9999 Monte-Carlo resampling with the "Wilcox_test" function in the R package "coin". The mean and standard deviation were calculated for three alpha-diversity measurements (OTU counts, phylogenetic diversity, and Shannon index) and relative abundances of four bacterial phyla and 17 bacterial genera described above for the four groups. The differences in bacterial measurements among the four groups were tested using Wilcoxon rank test with Bonferroni correction (Table S4).

4.2.8. Genome-wide association

To identify host genes that are associated with microbial measurements, Multivariate Linear Mixed Models was used for association test in GEMMA (version 0.94) using the entire exome (279,278 SNPs). A total of 21,953 genes were identified in our exome data (see Phifer-Rixey et al. *in review* for more details). Relatedness was accounted by using the relatedness matrix estimated in GEMMA. Population structure was accounted by using the first four genetic principle components as covariates calculated from SNPRelate (version 1.10.2). To control for bacteria that vary latitudinally (Thompson *et al.* 2017), we also used latitude as a covariate. To control for experimental and batch effects, we accounted for hidden factors by inferring 10 cofounders in PEER. Manhattan plot and QQ-plot were generated using R package "qqman". Phylogenetic diversity and 17 bacterial genera were chosen for genome-wide association analyses

based on average relative abundance of >1% across all samples representing four dominant phyla; Firmicutes (6 genera), Bacteroidetes (7 genera), Proteobacteria (3 genera), and Deferribacteres (1 genus). The motivation to only select 17 common bacterial genera is both statistical (e.g. avoid 0 values and minimize multiple testing) and biological reasons (e.g. host genetic control seems to be stronger at the tips of the phylogenetic tree (Benson *et al.* 2010a)). To control for false discovery, q-value was calculated using R package "qvalue" based on Likelihood ratio p-values. Significant SNPs were called at q-value < 0.1.

To test the overlap of bacteria-associated genes identified between this study and previously published human and mouse mapping studies, we compiled SNPs and genes that were associated with microbial measurements in seven human mGWAS studies. For human mGWAS, we collected genes and SNPs that are reported as significantly associated with microbial measurements in humans and tested them in mice. We converted positions to genes using biomart (GRCh37) and identified human-mouse orthologous based on Ensembl (Sep. 2017). Among the 555 unique human genes that we collected from seven studies, a total of 469 genes had human-mouse orthologous based on Ensembl (Sep. 2017). This resulted in 10,194 SNPs for human candidates in our exome. To test whether the overlap of candidate genes between groups was significant, we used a hypergeometric test using "phyper" function in R. We used the total number of possible human-mouse orthologous genes in our exome (19,100) based on Ensembl database (Sep 2017). To test whether the proportion of significant gene-bacteria associations was greater in the candidate gene set compared to all genes using the entire exome, we used a chi-square test with Yate's correction.

To test whether different genotypes have significantly different relative abundances of bacterial taxa among the top candidate genes from each bacterial measurement and among the candidate genes that overlapped between wild mouse and human mGWAS, we used Wilcoxon/Kruskal-Wallis tests in JMP 13.0. Residuals after covariate regression (i.e. Genetic PC1-4 and Latitude) on box-cox transformed relative abundances of bacterial taxa and alpha-diversity were used for all tests. For the 20 genes identified in mouse mGWAS (q-value < 0.1), we first tested genotypic differences on the microbial measurements using all individuals using ANOVA. Next, to confirm the associations are not driven by population structure, we tested whether the directions of genotype-bacteria association persists within populations using ANOVA (Table S5). We used a sign test to ask whether the directionality of genotype-bacteria associations within populations deviates significantly from the expected 50:50 ratio. Effect sizes for the 20 candidate genes in mouse mGWAS (q-value < 0.1) and 8 candidate genes (q-value < 0.2) that overlapped between mouse and human mGWAS were calculated using ANOVA by using microbial measurements controlled for covariates described above.

4.3. Results and Discussions

4.3.1. Host genetic distance and body size are associated with compositional variations in the gut microbiota

The gut microbial community, as measured by Bray-Curtis dissimilarity, was significantly different among the five populations of house mice along the east coast of North America (Fig. 1A&B, ADONIS, R^2=0.095, p=0.04) consistent with findings in European populations of house mice (26). However, the population differences observed in this study were relatively small. To understand how host genetic and environmental factors contribute to variation in gut microbial communities across all samples, we measured correlations between Bray-Curtis dissimilarity and eight predictor variables using Mantel tests (Table 1) and Generalized Linear Models (GLMs) (Table S1). Bray-Curtis dissimilarity was significantly correlated with both host genetic distance and body mass index (BMI) after correcting for multiple tests (Mantel r = 0.14, p = 0.004 and Mantel r = 0.25, p < 0.0001, respectively). Diet ($\delta^{13}C$), climate PC1, and geographic distance also showed weak correlations with Bray-Curtis dissimilarity (Table 1). Diet measurements ($\delta^{13}C$ and $\delta^{14}N$) did not vary among populations (Fig.S1) or by latitude (Table S2). GLMs that did not include host genetic distance, climate PC1, and body mass index (BMI) provided a significantly worse fit than a full model, indicating that these variables significantly correlate with Bray-Curtis dissimilarity after controlling for all other variables (Table S1). The overall results were similar using other beta-diversity measurements (Table S3).

The observed correlation between microbial distance and genetic distance was independent of geographic distance. First, there was no pattern of genetic isolation-by-distance among these populations (Fig. S2). Second, the correlation between host genetic distance and Bray-Curtis dissimilarity remained significant after controlling for geographic distance using a partial mantel test (r = 0.12, p = 0.008) and controlling for all other variables using GLMs (Table S1, likelihood ratio test p = 0.008). When comparisons were made between individuals within populations, both host genetic distance and geographic distance showed significant correlations with Bray-Curtis dissimilarity (Fig.1C&D). Consistent with the results among populations, the correlation between host genetic distance and Bray-Curtis dissimilarity within populations remained significant after controlling for geographic distance using residuals of covariate regression (Spearman's rho = 0.26, p < 0.0001). In contrast, the correlation between geographic distance and Bray-Curtis dissimilarity did not remain significant after controlling for genetic distance (Spearman's rho = 0.06, p = 0.38). These results suggest that host genetics or vertical transmission have stronger effects on the gut microbiota than the geographic distance between individuals.

The correlation between BMI and Bray-Curtis dissimilarity also remained significant after controlling for geographic distance (partial mantel r = 0.25, p < 0.001) and all other variables (Table S1, GLMs likelihood ratio test p = 0.03). This association is interesting because BMI and body weight vary clinally with latitude (Table S2), a pattern consistent with Bergmann's rule presumably reflecting thermoregulatory adaptation (Lynch 1992) (Phifer-Rixey et al. *in review*). Moreover, at higher latitudes, individuals tend to have a more obesity-associated microbial composition (i.e. a greater ratio of Firmicutes/Bacteroidetes) (Fig. S3). This result is consistent with observations in human populations (Suzuki and Worobey 2014). We also identified various microbial taxa that correlated with latitude and body size measurements after accounting for latitude (Fig. S4). Further experiments, including transplants into gnotobiotic mice, would be useful for

testing the role of the gut microbiome in Bergmann's rule and thermoregulatory adaptation.

4.3.2. Population differences in the microbiota persist in a common laboratory environment

The two populations from the ends of the transect had significantly different microbial compositions (Fig.1E). To test whether these differences were driven by environmental differences, we conducted a common garden experiment. Twenty unrelated wild mice were collected from each population and returned to the lab (10 males, 10 females). For each population, 10 crosses between wild-caught parents were created to produce 40 offspring which were reared under identical conditions. The lab-born mice showed major shifts in alpha-diversity and in the relative abundances of bacterial phyla and genera compared to the wild-caught animals (Table S4). For example, alpha-diversity and the relative abundances of Firmicutes and Proteobacteria significantly decreased, and the relative abundance of Bacteroidetes significantly increased in lab-reared animals compared to wild-caught animals (Table S4), consistent with previous studies (Wang *et al.* 2014, 2015). Interestingly, the lab diet was not significantly different from the diet in the wild as assessed from carbon and nitrogen stable isotope measurements (Fig. S1).

Despite the dramatic shifts in the microbiota from the wild to the lab environment, population differences in the microbiota persisted among lab-reared offspring (Fig.1E). Moreover, the microbial community composition of lab populations was more similar, on average, to the wild populations from which they came than to the wild populations at the other end of the transect (Wilcoxon permutation test, $P = 0.029$, Fig. S5). Although the gut microbiota of the southern wild population was equally similar to the gut microtiota of the northern and southern lab populations (Wilcoxon permutation test, $P = 0.44$), the gut microbiota of the northern wild population was significantly more similar to the gut microbiota of the northern lab population compared to the southern lab population (Wilcoxon permutation test, $P = 0.0014$). Overall, these results indicate that environmental differences alone (e.g. diet, environmental microbes, temperature, etc.) cannot explain the population differences in the microbiota. The observed population differences are consistent either with a role for host genetics or simply with vertical transmission shaping the variation of the gut microbiota in wild mice.

4.3.3. Identification of genetic loci underlying gut microbiota variation in wild mice

To identify host genes contributing to differences in the gut microbiota, we conducted a mGWAS using ~280,000 SNPs identified from sequencing the complete exomes of the 50 wild-caught mice (Phifer-Rixey et al. in review). We searched for associations between host genetic variation and each of 17 bacterial genera that had an average relative abundance of > 1%. We also searched for associations between host genetic variation and alpha-diversity represented by phylogenetic diversity. Analyses were done using Multivariate Linear Mixed Models in GEMMA while controlling for population structure, relatedness, latitude, and hidden factors. Among the 18 bacterial

measurements, three showed significant associations with host genetic loci; *Odoribacter*, *Bacteroides*, and phylogenetic diversity (Table 2). Across all tests, we identified a total of 24 SNPs in 20 genes that passed a genome-wide significance threshold (q-value < 0.1). Although none of the overrepresented GO terms were significant after false discovery correction, the top three GO terms include mRNA transcription (*Mier1*, p-value = 0.006), protein lipidation (*Zdhhc7,* p-value = 0.04), and nucleobase-containing compound transport (*Slc35d1*, p-value = 0.07).

Although we attempted to account for population structure in identifying these genes using GEMMA, observed gene-bacteria associations might still be be driven by differences among populations that are not fully accounted for by the model. To further account for population structure, we looked at associations within individual populations and we asked whether the direction of the association was consistent among populations. Overall, most of the within-population genotype-bacteria comparisons showed the same direction as the all-population comparisons (37 out of 43 comparisons, sign test p-value < 0.0001, Table S5). Moreover, 20 of these 37 comparisons were individually significant (ANOVA, p-value < 0.05) despite the fact that these tests are underpowered with only 10 individuals per population (Table S5). Together, the results suggest that the observed genotype-bacteria associations are unlikely to be explained by population structure.

Among the 20 genes that were associated with bacterial measurements, the interleukin 12a gene (*IL12a*) included a SNP with the lowest p-value across all tests in this study (Table 2). *IL12* is a cytokine that plays a key role in innate and adaptive immunity by activating natural killer cells and regulating differentiation of T cells (Trinchieri 1998). Six SNPs in *IL12a* were significantly associated with the relative abundance of *Odoribacter* after accounting for population structure, latitude, and hidden factors as covariates (Fig. 2A-D, Table 2). Three of these six SNPs were non-synonymous changes (Table 2). The up-regulation of *IL12a* production has been linked Crohn's disease in humans (Parronchi *et al.* 1997) and mucosal inflammation in mice (Liu *et al.* 2001). Furthermore, a recent study in humans demonstrated that inflammatory cytokine responses are associated with microbial taxa composition, metagenomic functional profiles, and microbial metabolites (Schirmer *et al.* 2016). Interestingly, the relative abundance of *Odoribacter* was significantly correlated with tumor necrosis factor alpha (TNF-α) (34), which is another macrophage-derived cytokine that interacts with *IL12a* in mediating inflammatory responses in mammals (Ma 2001). These observations lend further support to the role of *IL12a* in mediating host-microbial interactions in wild mice.

4.3.4. Homologous genes underlie gut microbiota variation in humans and mice

A common set of genes may underlie host-bacterial interactions across diverse mammals. To test this idea, we asked whether there was significant overlap between the genes underlying variation in the microbiota of mice and humans using two different approaches. First, we compiled genes that were associated with microbial measurements in seven different human mGWAS. This comprised a set of 469 genes with one-to-one mouse-human orthologs (Table S6). We then conducted association analyses in GEMMA using this set of 469 genes in mice and found that 10 were significantly associated with one or more bacterial measurements (q-value < 0.1) (Table S6). This fraction of genes

showing associations (10 out of 469 = 2.13%) is significantly greater than the fraction discovered in the initial analysis using all genes (20 out of 21,954 = 0.09%) suggesting that mouse mGWAS hits are overrepresented among genes previously identified in human mGWAS (Chi-square test with Yate's correction $p < 0.0001$).

Second, we asked how many genes overlapped between the 469 genes identified in human mGWAS and the 20 mouse-human orthologous genes that were identified in the mouse mGWAS. Using the genome-wide cut-off of q-value < 0.1, there was only one gene, *Csmd1* that overlapped between these sets, and this proportion of overlap was marginally not significant (hypergeometric test $p = 0.06$). However, when we made the genome-wide cut-off less stringent (q-value < 0.2), we identified 96 mouse-human orthologous genes (Table S7) and eight genes overlapped with 469 human candidate genes (Fig. 2E, Table S8). The number of overlapped genes identified at q-value < 0.2 was more than expected by chance (hypergeometric test $p = 0.0006$). Not surprisingly, all eight genes were the same genes identified by the candidate gene approach mentioned above (Table S6 & S8).

Among the eight homologous overlapping genes between human and mouse mGWAS, all show expression in the brain of mice and humans using public databases (Table S8) and some have been associated with phenotypes related to obesity and immunity in other human GWAS. For example, a SNP in *Csmd1* was associated with alpha-diversity (i.e. phylogenetic diversity) and showed the lowest p-value among the eight human-mouse overlapping genes. *Csmd1* is highly expressed in central nervous system and epithelial tissue and is involved in the regulation of development of the central nervous system (Kraus *et al.* 2006). In humans, *Csmd1* was associated with beta-diversity of the gut microbiota (Wang *et al.* 2016), obesity-related traits (Irvin *et al.* 2011; Comuzzie *et al.* 2012; Liu *et al.* 2013), parasite infection status (Deng *et al.* 2013), and antibody response to smallpox vaccine (Ovsyannikova *et al.* 2012). Similarly, *Gpr158* is also highly expressed in mouse and human brains and was associated with bacterial taxa in the order Clostridiales in both wild mice and humans (Goodrich *et al.* 2016a). *Gpr158* has been associated with variation in energy expenditure in a native American population that has a high prevalence of obesity (Piaggi *et al.* 2017). These results support the previously known links between gut microbial measurements and mammalian host genes related to the nervous system, immunity, and obesity.

4.3.5. Conclusion

Understanding how mammalian hosts control the composition and function of the gut microbiota remains a major challenge in microbial ecology and biomedical research. We showed that differences in host genome and body mass were associated with compositional differences in the wild mouse gut microbiota. We next identified population differences in the gut microbiota that persist in a common laboratory environment, supporting the role of host genetics or vertical transmission in shaping variation of the gut microbial communities. Finally, we identified both novel and previously known gene-bacteria associations in wild mice using a genome-wide mapping approach. Moreover, a significantly greater number of genes were associated with gut microbial variation in both humans and mice than expected by chance, including genes related to the nervous system, immunity, and obesity. Replicating these results in

independent populations of wild mice and validating the functions of candidate SNPs in wild-derived inbred mice would further strengthen the observed gene-bacteria associations. Gene-bacteria associations identified in wild mice and human subjects using similar mapping methods are strong candidates for genes influencing the mammalian gut microbial composition in their natural environment.

4.4. Chapter 4 Tables

Table 1. Correlations between microbial beta-diversity and predictor variables using Mantel test.

	n	Bray Curtis distance	
		Mantel r	p-value
Genetic distance	50	**0.14**	**0.004**
Geographic distance	50	0.07	0.078
Body weight	50	0.08	0.232
BMI	50	**0.25**	**<0.001**
Diet ($\delta^{13}C$)	50	0.14	0.044
Diet ($\delta^{15}N$)	50	0.04	0.67
ClimatePC1	50	0.10	0.066
ClimatePC2	50	0.01	0.913

*Bonferroni corrected p-value = 0.05/8 = 0.0063

Table 2. Loci significantly associated with microbial measurements (q-value < 0.1) in wild mouse mGWAS.

Chr	Bp	Annotated gene(s)	Associated microbial measurements	P-values[1]	Q-values	Effect size (%)[2]	Missense?
3	68695209	*Il12a*	*Odoribacter*	8.41E-07	0.042	23.1	yes
3	68695333			1.10E-06	0.045	19.9	yes
3	68695379			5.37E-07	0.033	18.0	-
3	68695382			5.37E-07	0.033	18.0	-
3	68695502			4.13E-07	0.033	22.2	-
3	68695548			7.79E-09	0.002	24.4	yes
3	86138475	*Snord73a, Rnu73b, Rps3a1*	*Bacteroides*	1.77E-06	0.054	18.1	-
3	86138574			6.90E-07	0.028	18.4	-
3	86138625			2.00E-06	0.054	29.8	-
4	103170679	*Mier1, Slc35d1*	Phylogenetic diversity	2.19E-06	0.092	5.4	-
4	89692441	*Dmrta1*	*Bacteroides*	4.73E-07	0.023	27.9	-
5	90490831	*Afp*	*Bacteroides*	1.27E-08	0.002	27.0	-
5	90490846			1.27E-08	0.002	27.0	-
5	90491657			8.67E-08	0.007	17.5	-
6	121221243	*Tuba8, Gm15856*	*Bacteroides*	4.19E-06	0.086	23.4	-
6	121222841			3.03E-06	0.074	23.3	-
6	128374454	*Foxm1, Tex52*	Phylogenetic diversity	4.09E-07	0.051	33.3	-
6	128374521			4.09E-07	0.051	33.3	-
6	128374742			9.63E-07	0.055	34.0	-
8	120092803	*Zdhhc7, Gm20388, Gm15898*	*Bacteroides*	3.38E-07	0.021	38.2	-
8	13142468	*Cul4a*	*Bacteroides*	3.56E-06	0.079	22.5	-
8	16358320	*Csmd1*	Phylogenetic diversity	1.09E-06	0.055	17.0	-
11	3132802	*Sfi1, Pisd-ps1*	Phylogenetic diversity	8.39E-07	0.055	12.0	-
13	33671503	*Serpinb6d*	*Bacteroides*	1.38E-06	0.048	31.7	-

[1] Likihood ratio p-values.

[2] ANOVA R^2 values. Residuals after covariate regression (i.e. Genetic PC1-4 and Latitude) on the box-cox transformed relative abundance of bacterial taxa was used.

4.5. Chapter 4 Figures

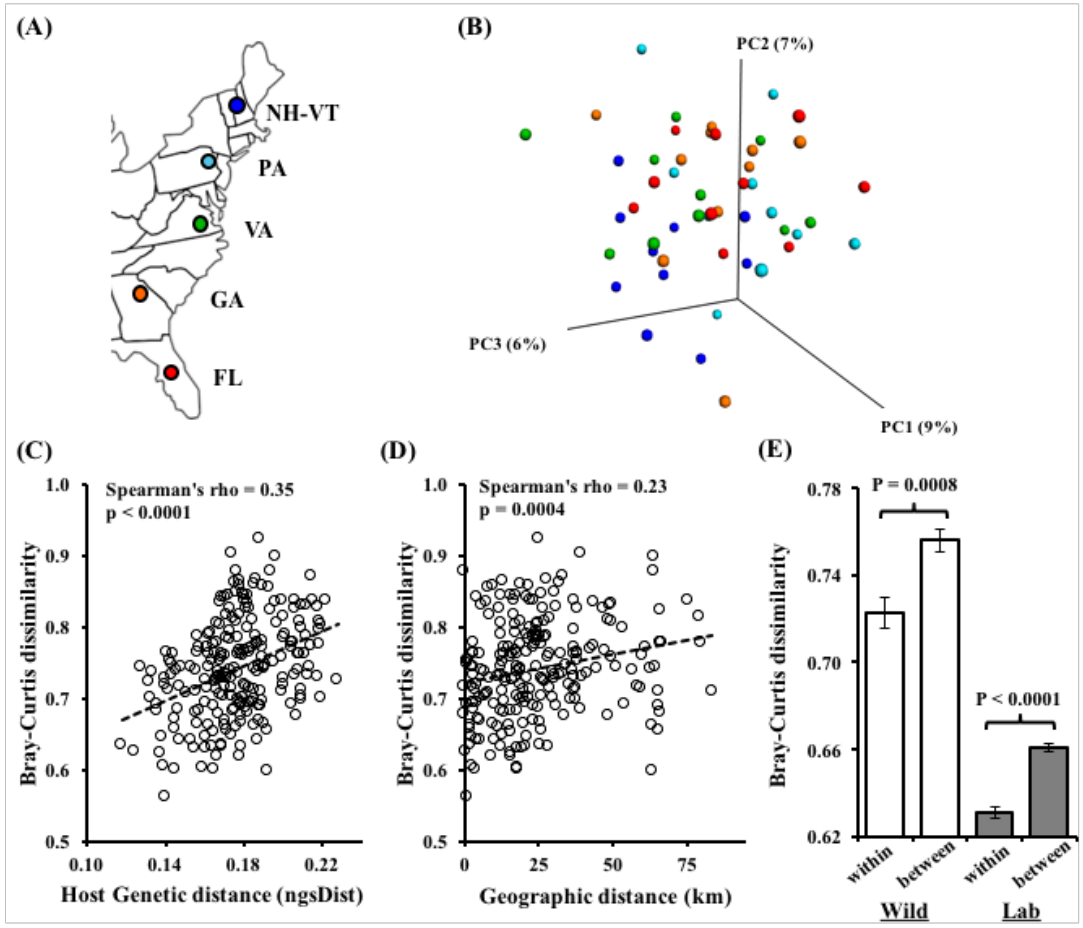


Figure 1. Population differences and host genetic distance are associated with compositional variations in the gut microbiota. (A) Sampling locations of five house mouse populations. (B) PCoA plot of Bray-Curtis dissimilarity. The color corresponds to populations in Fig.1A. Populations show weak, but significant clustering (ADONIS R^2=0.095, p=0.04). (C) Positive correlation between microbial distance and host genetic distance (rho = 0.35, P < 0.0001) and (D) between microbial distance and geographic distance (rho = 0.23, P = 0.0004) within populations. Correlations for all individual comparisons (including between population comparisons) are shown in Table 1. (E) Average Bray-Curtis dissimilarity within population comparisons (within) and between population comparisons (between) of most northern and southern populations in the wild (white bars) and lab (gray bars). P-values are Wilcoxon permutation test based on 9999 Monte-Carlo resampling. Error bars are SE.

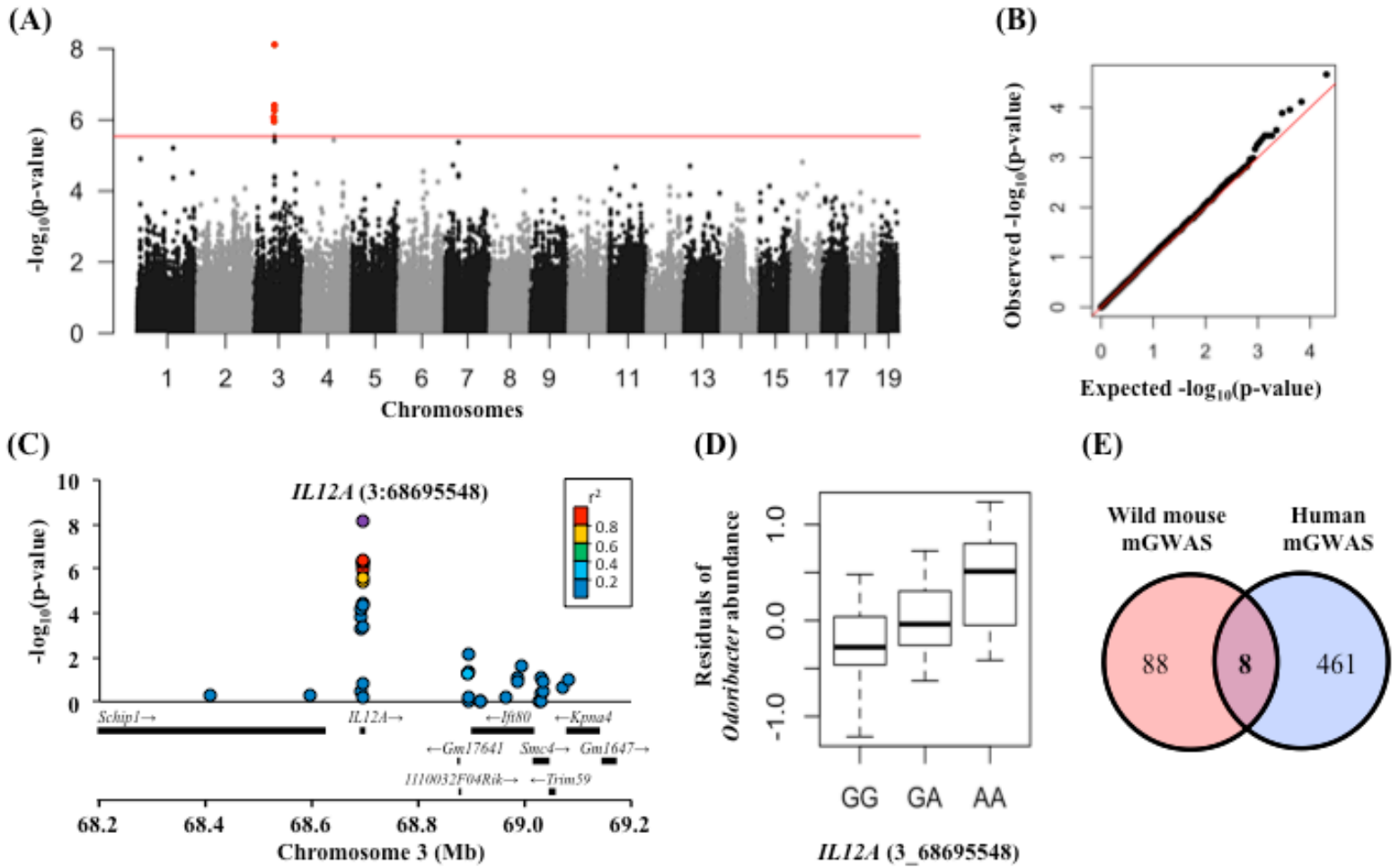


Figure 2. Results of mGWAS in wild mice. (A) Manhattan plot of *Odoribacter*. Six SNPs in *IL12A* gene on chromosome 3 are highlighted. Red line shows the genome-wide cut-off of the likelihood ratio test p-value (q-value < 0.1). (B) Quantile-quantile plot of *Odoribacter* p-values. The red diagonal line represents the expected distributions of p-values. (C) Zoom in plot around the SNP (3:68695548) that has the lowest p-value (represented by purple). Rest of the color represents pairwise linkage disequilibrium measures (r^2) between the SNP and SNPs within the surrounding 1 Mb window. r^2 was calculated based on genotype allele counts using PLINK. (D) Box-plot of *Odoribacter* abundance and *IL12A* genotypes. A missense SNP in *IL12A* gene has significantly different abundances of *Odoribacter* (ANOVA $R^2 = 0.224$, $p = 0.0014$). Residual values were used for the y-axis controlling for population structure, latitude, and hidden factors using covariate regression. (E) Eight genes were associated with mouse and human microbial measurements and this overlap is more than expected by chance (hypergeometric test $p = 0.0006$). A total of 96 mouse-human orthologous genes were identified in wild mouse using the cut-off of q-value < 0.2.

4.6. Supplemental information

Table S1. Model selection using Generalized Linear Models on Bray-Curtis dissimilarity PC1-3.

Generalized Linear Models[1]		Bray-Curtis PC1 (9%)		Bray-Curtis PC2 (7%)		Bray-Curtis PC3 (6%)	
	Fixed effects[2]	AICc	Log likelihood[3]	AICc	Log likelihood[3]	AICc	Log likelihood[3]
Full Model	Genetics + Geography + BW + BMI + Carbon + Nitrogen + ClimatePC1 + ClimatePC2	-29.88	27.83	-53.36	39.57	-54.45	40.12
Model 1 (no Genetics)	Geography + BW + BMI + Carbon + Nitrogen + ClimatePC1 + ClimatePC2	-32.03	27.32	-56.53	38.39	-50.62	36.62***
Model 2 (no Geography)	Genetics + BW + BMI + Carbon + Nitrogen + ClimatePC1 + ClimatePC2	-31.44	27.03	-54.93	38.77	-57.37	39.99
Model 3 (no BW)	Genetics + Geography + BMI + Carbon + Nitrogen + ClimatePC1 + ClimatePC2	-32.23	27.42	-55.98	39.30	-55.24	38.93
Model 4 (no BMI)	Genetics + Geography + BW + Carbon + Nitrogen + ClimatePC1 + ClimatePC2	-32.78	27.70	-51.66	37.14**	-57.59	40.10
Model 5 (no Carbon)	Genetics + Geography + BW + BMI + Nitrogen + ClimatePC1 + ClimatePC2	-30.30	26.46*	-56.51	39.56	-57.59	40.10
Model 6 (no Nitrogen)	Genetics + Geography + BW + BMI + Carbon +ClimatePC1 + ClimatePC2	-33.00	27.81	-56.35	39.48	-57.50	40.06
Model 7 (no ClimatePC1)	Genetics + Geography + BW + BMI + Carbon + Nitrogen + ClimatePC2	-30.50	26.56	-49.82	36.22***	-56.94	38.78
Model 8 (no ClimatePC2)	Genetics + Geography + BW + BMI + Carbon + Nitrogen + ClimatePC1	-30.37	26.49	-56.53	39.57	-56.21	39.41

[1] The response variable, Bray-Curtis dissimilarity PC1-3 was calculated using all wild individuals. The fractions of the variation explained were PC1 (9%), PC2 (7%), and PC3 (6%).

[2] Eight continuous variables were used for the Full Model. "Genetics" is an average genetic distance (ngsDist) from all individual comparisons. "Geography" is an average genetic distance (km) from all individual comparisons. "BW" is body weight (g). "BMI" is body mass index."Carbon" and "Nitrogen" are stable isotope diet measurements. "Climate PC1" and "Climate PC2" are climate PC axes. Each variable was substracted from the Full Model for model comparisons (Models 1-8). See SI methods for more detail.

[3] Significant differences compared to the full model were determined based on the likelihood ratio test; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table S2. Correlations with latitude and metadata.

Variable	All individuals			Only adults		
	n	rho	p-value	n	rho	p-value
Body weight	50	0.32	**0.02**	40	0.51	**0.0008**
BMI	50	0.21	0.14	40	0.35	**0.03**
Diet (Carbon)	49	0.19	0.19	40	0.25	0.12
Diet (Nitrogen)	49	0.14	0.35	40	0.24	0.14
Climate PC1	50	-0.97	**<0.0001**	40	-0.98	**<0.0001**
Climate PC2	50	0.32	**0.02**	40	0.23	0.15

Table S3. Correlations between microbial beta-diverisity measurements and predictor variables using Mantel test.

	n	Bray Curtis		unweighted UniFrac		weighted UniFrac	
		Mantel r	p-value*	Mantel r	p-value*	Mantel r	p-value*
Genetic distance	50	**0.14**	**0.004**	0.12	*0.012*	0.10	*0.037*
Geographic distance	50	0.07	0.078	0.10	*0.035*	0.06	0.148
Body weight	50	0.08	0.232	0.12	0.121	0.07	0.337
BMI	50	**0.25**	**<0.001**	0.12	0.165	**0.23**	**0.005**
Diet ($\delta^{13}C$)	50	0.14	0.044	0.06	0.482	-0.01	0.898
Diet ($\delta^{15}N$)	50	0.04	0.67	-0.01	0.924	0.00	0.993
ClimatePC1	50	0.10	0.066	0.09	0.156	0.10	0.072
ClimatePC2	50	0.01	0.913	0.04	0.488	-0.05	0.403

*Raw P-values that are lower than Bonferroni corrected p-value = 0.05/8 = 0.0063 are bolded

Table S4. Differences in microbial measurements among wild and lab populations.

Microbial measurements	Wild				Lab				Significance (Wilcoxon test)				
	North (n=10)		South (n=10)		North (n=40)		South (n=38)		Wild vs Lab[1]	Wild		Lab	
										North	South	North	South
	Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*	Raw p-values	Pairwise comparisons[2]			
Alpha-diversity													
OTU counts	408.5	*(46.8)*	351.2	*(41.4)*	365.7	*(43.5)*	345.9	*(44.7)*	NS	a	ab	ab	b
Phylogenetic diversity	38.7	*(3.4)*	34.6	*(2.5)*	33.7	*(3.2)*	32.6	*(3.4)*	0.0005	a	b	b	b
Shannon index	6.57	*(0.41)*	6.15	*(0.63)*	5.87	*(0.45)*	5.89	*(0.41)*	0.0002	a	ab	b	b
Phylum Firmicutes	0.477	*(0.100)*	0.426	*(0.162)*	0.284	*(0.090)*	0.366	*(0.150)*	0.0004	a	ab	b	ab
Genus *Clostridiales_unc*	0.251	*(0.081)*	0.213	*(0.120)*	0.181	*(0.054)*	0.234	*(0.101)*	NS	a	a	a	a
Genus *Lachnospiracea_unc*	0.063	*(0.040)*	0.057	*(0.030)*	0.025	*(0.017)*	0.044	*(0.028)*	0.0003	a	a	b	a
Genus *Lactobacillus*	0.016	*(0.023)*	0.007	*(0.006)*	0.014	*(0.016)*	0.022	*(0.018)*	0.01	a	a	a	a
Genus *Oscillospira*	0.062	*(0.023)*	0.064	*(0.014)*	0.028	*(0.016)*	0.031	*(0.020)*	<0.0001	a	a	b	b
Genus *Ruminococcaceae_unc*	0.049	*(0.014)*	0.050	*(0.023)*	0.018	*(0.013)*	0.021	*(0.019)*	<0.0001	a	a	b	b
Genus *Ruminococcus*	0.013	*(0.006)*	0.017	*(0.008)*	0.004	*(0.002)*	0.006	*(0.004)*	<0.0001	ac	a	b	bc
Phylum Bacteroidetes	0.325	*(0.109)*	0.420	*(0.201)*	0.618	*(0.100)*	0.535	*(0.158)*	<0.0001	a	ab	b	b
Genus *Bacteroides*	0.084	*(0.051)*	0.112	*(0.089)*	0.090	*(0.057)*	0.093	*(0.071)*	NS	a	a	a	a
Genus *Bacteroidales_unc*	0.036	*(0.022)*	0.016	*(0.020)*	0.073	*(0.035)*	0.064	*(0.043)*	<0.0001	ab	a	b	b
Genus *Odoribacter*	0.013	*(0.013)*	0.013	*(0.020)*	0.044	*(0.038)*	0.058	*(0.038)*	<0.0001	a	a	b	b
Genus *Parabacteroides*	0.016	*(0.015)*	0.012	*(0.007)*	0.003	*(0.006)*	0.004	*(0.004)*	<0.0001	a	a	b	b
Genus *Prevotella*	0.020	*(0.024)*	0.022	*(0.033)*	0.011	*(0.035)*	0.004	*(0.015)*	<0.0001	a	a	b	b
Genus *Rikenellaceae_unc*	0.046	*(0.029)*	0.033	*(0.026)*	0.181	*(0.092)*	0.148	*(0.083)*	<0.0001	a	a	b	b
Genus *S24_7_unc*	0.103	*(0.053)*	0.199	*(0.129)*	0.202	*(0.128)*	0.149	*(0.102)*	NS	a	a	a	a
Phylum Proteobacteria	0.161	*(0.054)*	0.117	*(0.103)*	0.072	*(0.037)*	0.063	*(0.036)*	<0.0001	a	ab	b	b
Genus *Desulfovibrionaceae_unc*	0.044	*(0.042)*	0.027	*(0.038)*	0.001	*(0.002)*	0.004	*(0.006)*	<0.0001	a	ab	b	b
Genus *Helicobacteraceae_unc*	0.039	*(0.025)*	0.047	*(0.102)*	0.050	*(0.034)*	0.029	*(0.024)*	NS	a	a	a	a
Genus *Helicobacter*	0.064	*(0.064)*	0.001	*(0.003)*	0.165	*(0.019)*	0.008	*(0.013)*	NS	ab	a	b	a
Phylum Deferribacteres	0.021	*(0.016)*	0.028	*(0.052)*	0.004	*(0.005)*	0.012	*(0.014)*	0.03	a	ab	b	ab
Genus *Mucispirillum*	0.021	*(0.016)*	0.028	*(0.052)*	0.004	*(0.005)*	0.012	*(0.014)*	0.03	a	ab	b	ab

[1] Comparisons between Wild (North and South combined, n=20) and Lab (North and South combined, n = 78).

[2] Non-overapping letters indicate significance based on Wilcoxon test using Bonferroni corrected p-value (alpha = 0.05/24 = 0.002)

Table S5. Genotype-bacteria associations using all populations and within populations.

Chr	Bp	Genes[1]	Bacterial measurements[2]	All populations		Within Populations					
				Genotype directions[3]	ANOVA P-values	Consistent genotype directions[4]	ANOVA P-values[5]				
							FL	GA	VA	PA	NH VT
3	68695548	*Il12a*	*Odoribacter*	AA > GA > GG	0.0014	3 out of 5	0.020	0.001	(0.259)	0.428	(0.045)
3	86138574	*Snord73a, Rnu73b, Rps3a1*	*Bacteroides*	AC > AA	0.0021	2 out of 3	0.013	-	-	0.001	(0.830)
4	103170679	*Mier1, Slc35d1*	Phylogenetic diversity	TT > TC	0.1107	2 out of 2	-	-	-	0.130	0.220
4	89692441	*Dmrta1*	*Bacteroides*	GG > GA > AA	0.0005	2 out of 3	0.017	-	(0.367)	< 0.0001	-
5	90490831	*Afp*	*Bacteroides*	AA > GA > GG	0.0006	3 out of 4	0.030	-	(0.336)	0.001	0.978
6	121222841	*Tuba8, Gm15856*	*Bacteroides*	TC > TT	0.0004	3 out of 4	0.010	-	0.209	0.001	(0.813)
6	128374454	*Foxm1, Tex52*	Phylogenetic diversity	GG > GT	< 0.0001	5 out of 5	0.046	0.018	0.128	0.025	0.250
8	120092803	*Zdhhc7, Gm20388, Gm15898*	*Bacteroides*	CT > CC	< 0.0001	3 out of 3	0.010	0.059	-	0.001	-
8	13142468	*Cul4a*	*Bacteroides*	CT > CC	0.0006	4 out of 4	0.334	0.076	-	0.001	0.720
8	16358320	*Csmd1*	Phylogenetic diversity	CC > CT > TT	0.0125	4 out of 4	0.681	0.352	-	0.014	0.541
11	3132802	*Sfi1, Pisd-ps1*	Phylogenetic diversity	CC > CA	0.0159	4 out of 4	0.262	0.560	-	0.041	0.220
13	33671503	*Serpinb6d*	*Bacteroides*	CT > CC	< 0.0001	2 out of 2	0.0014	-	-	0.0007	-
					Total:	37 out of 43					

[1] The SNP with lowest p-value was selected when more then one SNP was included in the gene.

[2] Residuals after covariate regression (i.e. Genetic PC1-4 and Latitude) on box-cox transformed relative abundances of bacterial taxa and alpha-diversity were used.

[3] The directions of the genotype-bacteria associations were based on the average values of bacterial measurements for each genotype.

[4] The proportions of individual populations that genotype-bacteria associations are in the the same direction as the all-population result.

[5] P-values in prentices indicate that genotype-bacteria associations are in the opposite direction compared to the all-population result.

Table S6. Loci significantly associated with microbial measurements (q-value < 0.1) in wild mouse mGWAS using human candidate genes.

Chr	Bp	Annotated gene(s)	Associated microbial measurements	P-values[1]	Q-values
2	21367906	*Gpr158*	*Clostridiales_unc*	9.03E-06	0.053
5	148315352	*Mtus2*	*Phylogenetic diversity*	5.44E-05	0.084
6	113765833	*Atp2b2*	*Ruminococcus*	7.55E-06	0.074
6	113817432		*Phylogenetic diversity*	4.01E-05	0.074
6	113817441		*Phylogenetic diversity*	1.31E-05	0.030
6	144993723	*Bcat1*	*Clostridiales_unc*	1.12E-05	0.053
6	48445226	*Sspo*	*Bacteroides*	4.38E-05	0.088
8	16358320	*Csmd1*	*Phylogenetic diversity*	1.09E-06	0.010
8	16358359			3.89E-06	0.018
12	104780917	*Clmn*	*Bacteroides*	3.07E-05	0.088
12	75308870	*Bcat1*	*Bacteroides*	1.90E-05	0.088
12	75308871			5.83E-05	0.093
12	75308874			1.53E-05	0.088
12	75394205	*Rhoj*	*Phylogenetic diversity*	5.70E-06	0.018
13	69612885	*Nsun2*	*Bacteroides*	4.40E-05	0.088
15	5217545	*Ptger4*	*Bacteroides*	6.50E-05	0.093

[1] Likelihood ratio p-values.

[2] ANOVA R^2 values. Residuals after covariate regression (i.e. Genetic PC1-4 and Latitude) on box-cox transformed relative abundances of bacterial taxa and alpha-diversity were used.

Table S7. Significant loci associated with microbial measurements in wild mouse mGWAS (Q<0.2)

Taxa	chr	ps	allele1	allele0	p_lrt	qvalues	p_score	EnsembleID
Phylogenetic diversity	6	128374454	T	G	4.09E-07	0.051478057	4.62E-05	ENSMUSG00000001517
Phylogenetic diversity	6	128374454	T	G	4.09E-07	0.051478057	4.62E-05	ENSMUSG00000001517
Phylogenetic diversity	6	128374454	T	G	4.09E-07	0.051478057	4.62E-05	ENSMUSG00000079304
Phylogenetic diversity	6	128374521	C	T	4.09E-07	0.051478057	4.62E-05	ENSMUSG00000001517
Phylogenetic diversity	6	128374521	C	T	4.09E-07	0.051478057	4.62E-05	ENSMUSG00000001517
Phylogenetic diversity	6	128374521	C	T	4.09E-07	0.051478057	4.62E-05	ENSMUSG00000079304
Phylogenetic diversity	11	3132802	A	C	8.39E-07	0.054876364	6.32E-05	ENSMUSG00000023764
Phylogenetic diversity	11	3132802	A	C	8.39E-07	0.054876364	6.32E-05	ENSMUSG00000023764
Phylogenetic diversity	11	3132802	A	C	8.39E-07	0.054876364	6.32E-05	ENSMUSG00000082286
Phylogenetic diversity	11	3132802	A	C	8.39E-07	0.054876364	6.32E-05	ENSMUSG00000023764
Phylogenetic diversity	6	128374742	A	T	9.63E-07	0.054876364	6.71E-05	ENSMUSG00000001517
Phylogenetic diversity	6	128374742	A	T	9.63E-07	0.054876364	6.71E-05	ENSMUSG00000001517
Phylogenetic diversity	6	128374742	A	T	9.63E-07	0.054876364	6.71E-05	ENSMUSG00000079304
Phylogenetic diversity	8	16358320	T	C	1.09E-06	0.054876364	7.10E-05	ENSMUSG00000060924
Phylogenetic diversity	4	103170679	C	T	2.19E-06	0.09188015	9.78E-05	ENSMUSG00000028522
Phylogenetic diversity	4	103170679	C	T	2.19E-06	0.09188015	9.78E-05	ENSMUSG00000028521
Phylogenetic diversity	8	13142468	T	C	3.48E-06	0.122401981	1.22E-04	ENSMUSG00000031446
Phylogenetic diversity	8	13142468	T	C	3.48E-06	0.122401981	1.22E-04	ENSMUSG00000031446
Phylogenetic diversity	8	16358359	C	A	3.89E-06	0.122401981	1.28E-04	ENSMUSG00000060924
Phylogenetic diversity	8	16358359	C	A	3.89E-06	0.122401981	1.28E-04	ENSMUSG00000060924
Phylogenetic diversity	12	75394205	T	C	5.70E-06	0.159426744	1.55E-04	ENSMUSG00000046768
Phylogenetic diversity	12	75394205	T	C	5.70E-06	0.159426744	1.55E-04	ENSMUSG00000046768
Phylogenetic diversity	1	88236753	T	C	7.16E-06	0.161824139	1.73E-04	ENSMUSG00000079429
Phylogenetic diversity	1	88236753	T	C	7.16E-06	0.161824139	1.73E-04	ENSMUSG00000079429
Phylogenetic diversity	1	88236753	T	C	7.16E-06	0.161824139	1.73E-04	ENSMUSG00000079429
Phylogenetic diversity	5	90526381	A	G	8.32E-06	0.161824139	1.87E-04	ENSMUSG00000029369
Phylogenetic diversity	11	57787336	A	G	6.91E-06	0.161824139	1.70E-04	ENSMUSG00000020520
Phylogenetic diversity	6	128375274	T	C	8.79E-06	0.161824139	1.92E-04	ENSMUSG00000001517
Phylogenetic diversity	6	128375274	T	C	8.79E-06	0.161824139	1.92E-04	ENSMUSG00000001517
Phylogenetic diversity	6	128375274	T	C	8.79E-06	0.161824139	1.92E-04	ENSMUSG00000079304
Phylogenetic diversity	2	94301405	T	A	1.16E-05	0.161824139	2.21E-04	ENSMUSG00000027194
Phylogenetic diversity	7	141426886	T	A	1.20E-05	0.161824139	2.25E-04	ENSMUSG00000019082
Phylogenetic diversity	7	141426886	T	A	1.20E-05	0.161824139	2.25E-04	ENSMUSG00000060240
Phylogenetic diversity	7	141426886	T	A	1.20E-05	0.161824139	2.25E-04	ENSMUSG00000060240
Phylogenetic diversity	6	128374511	T	A	1.22E-05	0.161824139	2.27E-04	ENSMUSG00000001517
Phylogenetic diversity	6	128374511	T	A	1.22E-05	0.161824139	2.27E-04	ENSMUSG00000001517
Phylogenetic diversity	6	128374511	T	A	1.22E-05	0.161824139	2.27E-04	ENSMUSG00000079304
Phylogenetic diversity	7	24925094	A	G	1.25E-05	0.161824139	2.30E-04	ENSMUSG00000040940
Phylogenetic diversity	7	24925094	A	G	1.25E-05	0.161824139	2.30E-04	ENSMUSG00000040940
Phylogenetic diversity	12	32205995	T	C	1.27E-05	0.161824139	2.32E-04	ENSMUSG00000020573
Phylogenetic diversity	12	32205995	T	C	1.27E-05	0.161824139	2.32E-04	ENSMUSG00000020573
Phylogenetic diversity	6	113817441	T	C	1.31E-05	0.161824139	2.35E-04	ENSMUSG00000030302
Phylogenetic diversity	3	10314899	G	A	1.32E-05	0.161824139	2.36E-04	ENSMUSG00000027531
Phylogenetic diversity	2	91674525	T	G	1.35E-05	0.161824139	2.39E-04	ENSMUSG00000027247
Phylogenetic diversity	2	91674525	T	G	1.35E-05	0.161824139	2.39E-04	ENSMUSG00000027244
Phylogenetic diversity	7	134270142	T	C	1.65E-05	0.188794829	2.65E-04	ENSMUSG00000030994
Clostridiales_unc	9	106071836	T	C	5.05E-07	0.123470422	5.06E-05	ENSMUSG00000032572
Clostridiales_unc	2	21367906	T	C	9.03E-06	0.157699846	1.95E-04	ENSMUSG00000085680
Clostridiales_unc	2	21367906	T	C	9.03E-06	0.157699846	1.95E-04	ENSMUSG00000045967
Clostridiales_unc	2	21367906	T	C	9.03E-06	0.157699846	1.95E-04	ENSMUSG00000045967
Clostridiales_unc	3	87440123	G	T	7.48E-06	0.157699846	1.77E-04	ENSMUSG00000048031
Clostridiales_unc	3	87440335	C	G	5.25E-06	0.157699846	1.48E-04	ENSMUSG00000048031
Clostridiales_unc	3	87440340	T	C	9.02E-06	0.157699846	1.95E-04	ENSMUSG00000048031
Clostridiales_unc	3	87440341	G	A	9.02E-06	0.157699846	1.95E-04	ENSMUSG00000048031
Clostridiales_unc	3	87442103	A	G	2.84E-06	0.157699846	1.10E-04	ENSMUSG00000048031
Clostridiales_unc	3	87442103	A	G	2.84E-06	0.157699846	1.10E-04	ENSMUSG00000048031
Clostridiales_unc	3	87442187	G	A	2.84E-06	0.157699846	1.10E-04	ENSMUSG00000048031
Clostridiales_unc	3	87442187	G	A	2.84E-06	0.157699846	1.10E-04	ENSMUSG00000048031
Clostridiales_unc	3	87443684	A	C	6.14E-06	0.157699846	1.61E-04	ENSMUSG00000048031

Table S7. Continued.

Taxa	chr	ps	allele1	allele0	p_lrt	qvalues	p_score	EnsembleID
Clostridiales_unc	4	132356533	A	G	5.04E-06	0.157699846	1.46E-04	ENSMUSG00000066043
Clostridiales_unc	4	132356533	A	G	5.04E-06	0.157699846	1.46E-04	ENSMUSG00000085241
Clostridiales_unc	4	132356533	A	G	5.04E-06	0.157699846	1.46E-04	ENSMUSG00000066043
Clostridiales_unc	4	132356533	A	G	5.04E-06	0.157699846	1.46E-04	ENSMUSG00000091021
Clostridiales_unc	4	132356533	A	G	5.04E-06	0.157699846	1.46E-04	ENSMUSG00000064387
Clostridiales_unc	4	132356533	A	G	5.04E-06	0.157699846	1.46E-04	ENSMUSG00000065353
Clostridiales_unc	4	132356533	A	G	5.04E-06	0.157699846	1.46E-04	ENSMUSG00000028896
Clostridiales_unc	10	13553200	G	T	4.97E-06	0.157699846	1.45E-04	ENSMUSG00000019809
Clostridiales_unc	10	13553200	G	T	4.97E-06	0.157699846	1.45E-04	ENSMUSG00000019808
Clostridiales_unc	10	70448825	A	T	6.61E-06	0.157699846	1.67E-04	ENSMUSG00000043259
Clostridiales_unc	12	112775647	T	A	5.24E-06	0.157699846	1.48E-04	ENSMUSG00000072812
Clostridiales_unc	12	112775647	T	A	5.24E-06	0.157699846	1.48E-04	ENSMUSG00000072812
Clostridiales_unc	12	112775647	T	A	5.24E-06	0.157699846	1.48E-04	ENSMUSG00000072812
Clostridiales_unc	15	73783540	A	G	1.59E-06	0.157699846	8.42E-05	ENSMUSG00000072487
Clostridiales_unc	15	73783540	A	G	1.59E-06	0.157699846	8.42E-05	ENSMUSG00000072487
Clostridiales_unc	15	73783540	A	G	1.59E-06	0.157699846	8.42E-05	ENSMUSG00000072487
Clostridiales_unc	2	181601302	T	C	1.20E-05	0.162139377	2.25E-04	ENSMUSG00000038605
Clostridiales_unc	2	181601302	T	C	1.20E-05	0.162139377	2.25E-04	ENSMUSG00000002455
Clostridiales_unc	2	181601302	T	C	1.20E-05	0.162139377	2.25E-04	ENSMUSG00000002455
Clostridiales_unc	3	87443666	C	A	1.26E-05	0.162139377	2.31E-04	ENSMUSG00000048031
Clostridiales_unc	6	144993723	A	G	1.12E-05	0.162139377	2.17E-04	ENSMUSG00000030268
Clostridiales_unc	8	110971625	T	C	1.26E-05	0.162139377	2.30E-04	ENSMUSG00000015023
Clostridiales_unc	8	110971625	T	C	1.26E-05	0.162139377	2.30E-04	ENSMUSG00000031749
Clostridiales_unc	8	110971625	T	C	1.26E-05	0.162139377	2.30E-04	ENSMUSG00000031749
Clostridiales_unc	19	10218949	C	A	1.17E-05	0.162139377	2.22E-04	ENSMUSG00000036098
Clostridiales_unc	17	34216100	T	C	1.37E-05	0.167479682	2.41E-04	ENSMUSG00000024339
Clostridiales_unc	17	34216100	T	C	1.37E-05	0.167479682	2.41E-04	ENSMUSG00000024339
Clostridiales_unc	17	34216100	T	C	1.37E-05	0.167479682	2.41E-04	ENSMUSG00000081512
Bacteroides	5	90490831	A	G	1.27E-08	0.001556323	1.20E-05	ENSMUSG00000054932
Bacteroides	5	90490846	T	C	1.27E-08	0.001556323	1.20E-05	ENSMUSG00000054932
Bacteroides	5	90491657	A	C	8.67E-08	0.007083107	2.46E-05	ENSMUSG00000054932
Bacteroides	8	120092803	T	C	3.38E-07	0.020710124	4.27E-05	ENSMUSG00000089742
Bacteroides	8	120092803	T	C	3.38E-07	0.020710124	4.27E-05	ENSMUSG00000031823
Bacteroides	8	120092803	T	C	3.38E-07	0.020710124	4.27E-05	ENSMUSG00000031823
Bacteroides	8	120092803	T	C	3.38E-07	0.020710124	4.27E-05	ENSMUSG00000031823
Bacteroides	8	120092803	T	C	3.38E-07	0.020710124	4.27E-05	ENSMUSG00000092329
Bacteroides	4	89692441	G	A	4.73E-07	0.023185535	4.92E-05	ENSMUSG00000043753
Bacteroides	4	89692441	G	A	4.73E-07	0.023185535	4.92E-05	ENSMUSG00000043753
Bacteroides	3	86138574	C	A	6.90E-07	0.028185376	5.79E-05	ENSMUSG00000064984
Bacteroides	3	86138574	C	A	6.90E-07	0.028185376	5.79E-05	ENSMUSG00000064390
Bacteroides	3	86138574	C	A	6.90E-07	0.028185376	5.79E-05	ENSMUSG00000028081
Bacteroides	13	33671503	T	C	1.38E-06	0.048317787	7.89E-05	ENSMUSG00000047889
Bacteroides	3	86138475	C	T	1.77E-06	0.054226212	8.84E-05	ENSMUSG00000064984
Bacteroides	3	86138475	C	T	1.77E-06	0.054226212	8.84E-05	ENSMUSG00000064390
Bacteroides	3	86138475	C	T	1.77E-06	0.054226212	8.84E-05	ENSMUSG00000028081
Bacteroides	3	86138625	C	T	2.00E-06	0.054464494	9.36E-05	ENSMUSG00000064984
Bacteroides	3	86138625	C	T	2.00E-06	0.054464494	9.36E-05	ENSMUSG00000064390
Bacteroides	3	86138625	C	T	2.00E-06	0.054464494	9.36E-05	ENSMUSG00000028081
Bacteroides	6	121222841	C	T	3.03E-06	0.074262338	1.14E-04	ENSMUSG00000030137
Bacteroides	6	121222841	C	T	3.03E-06	0.074262338	1.14E-04	ENSMUSG00000086527
Bacteroides	8	13142468	T	C	3.56E-06	0.079320109	1.23E-04	ENSMUSG00000031446
Bacteroides	8	13142468	T	C	3.56E-06	0.079320109	1.23E-04	ENSMUSG00000031446
Bacteroides	6	121221243	A	G	4.19E-06	0.085577337	1.33E-04	ENSMUSG00000030137
Bacteroides	6	121221243	A	G	4.19E-06	0.085577337	1.33E-04	ENSMUSG00000086527
Bacteroides	5	90491714	T	C	5.90E-06	0.111233256	1.57E-04	ENSMUSG00000054932
Bacteroides	5	141952954	T	C	8.80E-06	0.154056712	1.92E-04	ENSMUSG00000039683
Bacteroides	2	91674525	T	G	9.96E-06	0.159308646	2.05E-04	ENSMUSG00000027247
Bacteroides	2	91674525	T	G	9.96E-06	0.159308646	2.05E-04	ENSMUSG00000027244
Bacteroides	4	148006547	T	C	1.04E-05	0.159308646	2.09E-04	ENSMUSG00000029016
Bacteroides	4	148006547	T	C	1.04E-05	0.159308646	2.09E-04	ENSMUSG00000029016

Table S7. Continued.

Taxa	chr	ps	allele1	allele0	p_lrt	qvalues	p_score	EnsembleID
Bacteroides	4	148006547	T	C	1.04E-05	0.159308646	2.09E-04	ENSMUSG00000086806
Bacteroides	4	148006547	T	C	1.04E-05	0.159308646	2.09E-04	ENSMUSG00000041616
Bacteroides	3	86139069	A	G	1.90E-05	0.167849669	2.86E-04	ENSMUSG00000064984
Bacteroides	3	86139069	A	G	1.90E-05	0.167849669	2.86E-04	ENSMUSG00000064390
Bacteroides	3	86139069	A	G	1.90E-05	0.167849669	2.86E-04	ENSMUSG00000028081
Bacteroides	4	89688171	T	C	1.18E-05	0.167849669	2.23E-04	ENSMUSG00000043753
Bacteroides	4	89688171	T	C	1.18E-05	0.167849669	2.23E-04	ENSMUSG00000043753
Bacteroides	5	90519074	C	A	2.01E-05	0.167849669	2.95E-04	ENSMUSG00000029369
Bacteroides	5	90526381	A	G	1.81E-05	0.167849669	2.79E-04	ENSMUSG00000029369
Bacteroides	5	118265460	G	A	2.09E-05	0.167849669	3.01E-04	ENSMUSG00000032840
Bacteroides	5	118265460	G	A	2.09E-05	0.167849669	3.01E-04	ENSMUSG00000095477
Bacteroides	6	41313290	T	G	2.26E-05	0.167849669	3.14E-04	ENSMUSG00000036938
Bacteroides	6	41313298	C	T	2.26E-05	0.167849669	3.14E-04	ENSMUSG00000036938
Bacteroides	6	41313339	C	T	2.26E-05	0.167849669	3.14E-04	ENSMUSG00000036938
Bacteroides	6	128374511	T	A	1.85E-05	0.167849669	2.83E-04	ENSMUSG00000001517
Bacteroides	6	128374511	T	A	1.85E-05	0.167849669	2.83E-04	ENSMUSG00000001517
Bacteroides	6	128374511	T	A	1.85E-05	0.167849669	2.83E-04	ENSMUSG00000079304
Bacteroides	7	46072425	T	C	2.11E-05	0.167849669	3.03E-04	ENSMUSG00000030835
Bacteroides	7	46081380	T	C	2.11E-05	0.167849669	3.03E-04	ENSMUSG00000030835
Bacteroides	7	46081380	T	C	2.11E-05	0.167849669	3.03E-04	ENSMUSG00000030835
Bacteroides	8	111715660	G	A	1.85E-05	0.167849669	2.82E-04	ENSMUSG00000031955
Bacteroides	12	75308870	G	A	1.90E-05	0.167849669	2.86E-04	ENSMUSG00000046768
Bacteroides	12	75308870	G	A	1.90E-05	0.167849669	2.86E-04	ENSMUSG00000046768
Bacteroides	12	75308870	G	A	1.90E-05	0.167849669	2.86E-04	ENSMUSG00000046768
Bacteroides	12	75308874	C	G	1.53E-05	0.167849669	2.56E-04	ENSMUSG00000046768
Bacteroides	12	75308874	C	G	1.53E-05	0.167849669	2.56E-04	ENSMUSG00000046768
Bacteroides	12	75308874	C	G	1.53E-05	0.167849669	2.56E-04	ENSMUSG00000046768
Bacteroides	13	55403064	A	G	1.46E-05	0.167849669	2.49E-04	ENSMUSG00000021490
Bacteroides	14	32598040	C	T	1.58E-05	0.167849669	2.60E-04	ENSMUSG00000041730
Bacteroides	14	70604927	A	G	1.47E-05	0.167849669	2.50E-04	ENSMUSG00000022099
Bacteroides	4	129472127	T	C	2.91E-05	0.17212018	3.60E-04	ENSMUSG00000040859
Bacteroides	4	129472127	T	C	2.91E-05	0.17212018	3.60E-04	ENSMUSG00000040859
Bacteroides	5	139771573	G	A	2.81E-05	0.17212018	3.53E-04	ENSMUSG00000029547
Bacteroides	5	139771573	G	A	2.81E-05	0.17212018	3.53E-04	ENSMUSG00000098574
Bacteroides	6	41216082	T	C	3.09E-05	0.17212018	3.72E-04	ENSMUSG00000076478
Bacteroides	6	41216085	T	C	3.09E-05	0.17212018	3.72E-04	ENSMUSG00000076478
Bacteroides	6	41303204	G	A	3.09E-05	0.17212018	3.72E-04	ENSMUSG00000054106
Bacteroides	6	41312455	G	A	3.09E-05	0.17212018	3.72E-04	ENSMUSG00000036938
Bacteroides	6	41312456	T	C	3.09E-05	0.17212018	3.72E-04	ENSMUSG00000036938
Bacteroides	6	41312497	T	A	3.09E-05	0.17212018	3.72E-04	ENSMUSG00000036938
Bacteroides	6	53816073	A	T	3.02E-05	0.17212018	3.67E-04	ENSMUSG00000078169
Bacteroides	6	53816073	A	T	3.02E-05	0.17212018	3.67E-04	ENSMUSG00000043496
Bacteroides	12	104780917	T	C	3.07E-05	0.17212018	3.70E-04	ENSMUSG00000021097
Bacteroides	19	4712684	A	G	2.93E-05	0.17212018	3.61E-04	ENSMUSG00000067889
Bacteroides	1	171237117	T	C	3.89E-05	0.198745891	4.22E-04	ENSMUSG00000058715
Bacteroides	1	171237117	T	C	3.89E-05	0.198745891	4.22E-04	ENSMUSG00000013593
Bacteroides	1	171237117	T	C	3.89E-05	0.198745891	4.22E-04	ENSMUSG00000013593
Bacteroides	1	171237117	T	C	3.89E-05	0.198745891	4.22E-04	ENSMUSG00000013593
Bacteroides	6	116692848	C	G	3.92E-05	0.198745891	4.24E-04	ENSMUSG00000048108
Bacteroides	6	116692848	C	G	3.92E-05	0.198745891	4.24E-04	ENSMUSG00000048108
Bacteroides	7	127375830	A	G	4.06E-05	0.198745891	4.32E-04	ENSMUSG00000054381
Bacteroides	9	72856998	A	G	4.09E-05	0.198745891	4.34E-04	ENSMUSG00000036030
Bacteroides	2	71209672	T	C	4.27E-05	0.198745891	4.44E-04	ENSMUSG00000027012
Bacteroides	6	48445226	A	C	4.38E-05	0.198745891	4.51E-04	ENSMUSG00000068551
Bacteroides	6	48445226	A	C	4.38E-05	0.198745891	4.51E-04	ENSMUSG00000029797
Bacteroides	6	48445226	A	C	4.38E-05	0.198745891	4.51E-04	ENSMUSG00000068551
Bacteroides	2	13010347	G	T	4.40E-05	0.198745891	4.51E-04	ENSMUSG00000049630
Bacteroides	13	69612885	T	C	4.40E-05	0.198745891	4.52E-04	ENSMUSG00000021594
Bacteroides	13	69612885	T	C	4.40E-05	0.198745891	4.52E-04	ENSMUSG00000091133
Bacteroides	13	69612885	T	C	4.40E-05	0.198745891	4.52E-04	ENSMUSG00000021595

Table S7. Continued.

Taxa	chr	ps	allele1	allele0	p_lrt	qvalues	p_score	EnsembleID
Bacteroides	10	40349413	C	A	4.46E-05	0.198745891	4.55E-04	ENSMUSG00000038491
Bacteroides	10	40349413	C	A	4.46E-05	0.198745891	4.55E-04	ENSMUSG00000038481
Bacteroides	10	40349416	A	C	4.46E-05	0.198745891	4.55E-04	ENSMUSG00000038491
Bacteroides	10	40349416	A	C	4.46E-05	0.198745891	4.55E-04	ENSMUSG00000038481
Bacteroides	14	79182968	T	C	4.46E-05	0.198745891	4.55E-04	ENSMUSG00000058997
Desulfovibrionaceae_unc	6	41617399	G	A	8.04E-07	0.137653188	6.20E-05	ENSMUSG00000029868
Desulfovibrionaceae_unc	6	41617399	G	A	8.04E-07	0.137653188	6.20E-05	ENSMUSG00000029869
Desulfovibrionaceae_unc	6	41617399	G	A	8.04E-07	0.137653188	6.20E-05	ENSMUSG00000029869
Desulfovibrionaceae_unc	6	41617399	G	A	8.04E-07	0.137653188	6.20E-05	ENSMUSG00000029869
Desulfovibrionaceae_unc	4	116557950	A	G	1.59E-06	0.137653188	8.42E-05	ENSMUSG00000034042
Desulfovibrionaceae_unc	4	116557950	A	G	1.59E-06	0.137653188	8.42E-05	ENSMUSG00000055900
Desulfovibrionaceae_unc	4	116557950	A	G	1.59E-06	0.137653188	8.42E-05	ENSMUSG00000034042
Desulfovibrionaceae_unc	4	88807526	T	C	1.69E-06	0.137653188	8.66E-05	ENSMUSG00000094648
Desulfovibrionaceae_unc	4	88807526	T	C	1.69E-06	0.137653188	8.66E-05	ENSMUSG00000070908
Desulfovibrionaceae_unc	4	88807202	G	T	3.60E-06	0.19206285	1.24E-04	ENSMUSG00000094648
Desulfovibrionaceae_unc	4	88807202	G	T	3.60E-06	0.19206285	1.24E-04	ENSMUSG00000070908
Desulfovibrionaceae_unc	16	17209775	G	C	3.93E-06	0.19206285	1.29E-04	ENSMUSG00000096434
Desulfovibrionaceae_unc	16	17209775	G	C	3.93E-06	0.19206285	1.29E-04	ENSMUSG00000071636
Odoribacter	3	68695548	A	G	7.79E-09	0.001924369	1.02E-05	ENSMUSG00000027776
Odoribacter	3	68695502	T	C	4.13E-07	0.033163867	4.64E-05	ENSMUSG00000027776
Odoribacter	3	68695379	A	G	5.37E-07	0.033163867	5.19E-05	ENSMUSG00000027776
Odoribacter	3	68695382	T	C	5.37E-07	0.033163867	5.19E-05	ENSMUSG00000027776
Odoribacter	3	68695209	G	C	8.41E-07	0.041550558	6.32E-05	ENSMUSG00000027776
Odoribacter	3	68695333	T	C	1.10E-06	0.045288956	7.11E-05	ENSMUSG00000027776

Table S8. Loci associated with microbial measurements in mouse and human mGWAS.

Chr	Bp	Annotated gene	Associated microbial measurements	P-values[1]	Q-values[2]	Effect size (%)[3]	Missense?	Tissue expressed[4]			
								Brain		Gut	
								mouse	human	mouse	human
2	21367906	*Gpr158*	*Clostridiales_unc*	9.03E-06	0.158	27.2	-	yes	yes	yes	-
6	48445226	*Sspo*	*Bacteroides*	4.38E-05	0.199	23.8	-	yes	NA	-	NA
6	113817441	*Atp2b2*	Phylogenetic diversity	1.31E-05	0.162	21.1	-	yes	yes	yes	-
6	144993723	*Bcat1*	*Clostridiales_unc*	1.12E-05	0.162	26.1	-	yes	yes	yes	yes
8	16358320	*Csmd1*	Phylogenetic diversity	1.09E-06	0.055	17.0	-	yes	yes	-	-
8	16358359			3.89E-06	0.122	13.8	-	yes	yes	-	-
12	75308870	*Rhoj*	*Bacteroides*	1.90E-05	0.168	28.7	yes	yes	yes	-	yes
12	75308874			1.53E-05	0.168	24.2	yes	yes	yes	-	yes
12	75394205	*Rhoj*	Phylogenetic diversity	5.70E-06	0.159	5.6	-	yes	yes	-	yes
12	104780917	*Clmn*	*Bacteroides*	3.07E-05	0.172	28.5	-	yes	yes	yes	yes
13	69612885	*Nsun2*	*Bacteroides*	4.40E-05	0.199	27.5	-	yes	yes	-	yes

[1] Likelihood ratio test p-values.

[2] SNPs with q-value < 0.2 were used to compare with human mGWAS gene set.

[3] ANOVA R^2 values. Residuals after covariate regression (i.e. Genetic PC1-4 and Latitude) on box-cox transformed relative abundances of bacterial taxa were used.

[4] Tissue expression data came from the Human Protein Atlas Database for humans and MGI Gene Expression Database for mice. All expression detected in brain tissue and gut (i.e. small and large intestines) as evidence of gene expression. "-" indicates no evidence of expression. "NA" indicates expression not tested.

Table S9. Metadata of wild-caught individuals.

ID	Localities	Latitude	Longitude	Sex	Body length	Weight	BMI	Repro. status	Carbon	Nitrogen	ClimatePC1	ClimatePC2	BarcodeSequence
MPR.108	FL	29.67445	-82.33093333	F	77	13	0.0022	adult	-21.8	5.2	2.9706	1.4993	CGGACTCGTTAC
MPR.110	FL	29.6233	-82.340333	F	90	23.5	0.0029	pregnant	-21.8	7	3.0965	1.5137	TCTCGCACTGGA
MPR.112	FL	29.778929	-82.416971	M	77	12.75	0.0022	adult	-18.8	5.9	2.9989	1.5015	TTCTGGTCTTGT
MPR.113	FL	29.78685	-82.495877	F	84	11.375	0.0016	adult	-20	6.2	2.9989	1.5015	GTCCACTTGGAC
MPR.114	FL	29.37283	-82.19892	F	84	15.5	0.0022	pregnant	-22.2	8	3.1006	1.4647	GATTTAGAGGCT
MPR.115	FL	29.45224	-82.3403	F	84	14	0.0020	adult	-22.6	4.5	3.2059	1.4694	GTCAGCCGTTAA
MPR.116	FL	29.39018	-82.1299	M	82	14.25	0.0021	adult	-17.9	7.1	3.0286	1.3517	ACGGTTTCTGGA
MPR.118	FL	29.10066	-82.14192	M	68	9.5	0.0021	adult	-21.8	7.1	3.1336	1.1473	GCAGCCATATTG
MPR.120	FL	29.7828	-82.3755	M	82.5	14	0.0021	adult	-24.2	7.7	2.9989	1.5015	ATAGGTGTGCTA
MPR.121	FL	29.65652	-82.5872	F	92	23	0.0027	pregnant	-16.8	7.6	3.0527	1.3108	ACCTAGCTAGTG
MPR.123	GA	34.0989	-83.3454	M	75.5	11	0.0019	adult	-17.9	7.1	0.6944	2.8708	GTCCTGACACTG
MPR.124	GA	34.10006	-83.59695	F	78	14.5	0.0024	pregnant	-19.9	6.5	0.7207	2.9541	GGACTCAACTAA
MPR.125	GA	34.09889	-83.34537	M	69	11	0.0023	adult	-21.4	9.9	0.6944	2.8708	ATACGGGTTCGT
MPR.126	GA	34.11877	-83.57092	M	80	18.5	0.0029	adult	-14.8	7.3	0.7207	2.9541	CCTTTCACCTGT
MPR.128	GA	33.87025	-83.28893	M	81	15.75	0.0024	adult	-21.3	7.8	0.5581	2.6117	ATCAGCCAGCTC
MPR.129	GA	33.83757	-83.35049	M	77	12.75	0.0022	adult	-20.4	6.6	0.5868	2.6940	GCTCCACAACGT
MPR.130	GA	33.94638	-83.38383	F	80	16.125	0.0025	adult	-17.3	6.4	0.5868	2.6940	AAGGAGTGCGCA
MPR.131	GA	33.88584	-83.48023	F	65	9	0.0021	pregnant	-9.5	9.6	0.5868	2.6940	AGGGAAAGGATC
MPR.132	GA	33.8437	-83.33744	F	75.5	11	0.0019	adult	-18.1	7.8	0.5868	2.6940	ACGACGCATTTG
MPR.133	GA	33.94639	-83.38384	F	89	21.375	0.0027	pregnant	-15.1	7.8	0.5868	2.6940	CGTCACTCCAAG
MPR.134	NH-VT	44.11202	-72.04543	F	83	19.125	0.0028	adult	-18	8.3	-2.9415	2.0775	TTACACAAAGGC
MPR.135	NH-VT	44.11212	-72.04551	F	93	21.375	0.0025	adult	-15.5	7.5	-2.9415	2.0775	GTATAGTCCGTG
MPR.137	NH-VT	43.99797	-72.10448	F	81	13.375	0.0020	adult	-17.6	10.8	-2.8739	2.4498	TCGTAAGCCGTC
MPR.138	NH-VT	43.97192	-72.08463	M	86	14.75	0.0020	adult	-23.4	5.7	-2.8739	2.4498	TGACGCCTCCAA
MPR.140	NH-VT	43.97281	-72.09393	M	74	14	0.0026	adult	-17.8	6.3	-2.8739	2.4498	TTCTCGGTTCTC
MPR.141	NH-VT	43.95075	-72.11753	M	86	19.5	0.0026	adult	-19.5	11.8	-2.8739	2.4498	GCTACTGGTATG
MPR.142	NH-VT	43.9609	-72.11585	F	94	23.75	0.0027	pregnant	-19.9	7.6	-2.8739	2.4498	GAATCCTCACCG
MPR.143	NH-VT	43.927907	-72.12301	M	81.5	15.5	0.0023	adult	-14.4	11.3	-2.8739	2.4498	CCTGACACACAC
MPR.144	NH-VT	44.12806	-72.04328	M	87.5	18.25	0.0024	adult	-15.5	7.5	-2.9415	2.0775	CAGCGTTTAGCC
MPR.145	NH-VT	43.95596	-72.11164	F	98	21	0.0022	adult	-19.3	6.7	-2.8739	2.4498	GGTATGGCTACT
MPR.146	PA	40.618763	-75.36362	F	86	13.125	0.0018	adult	-19.9	8	-1.2138	3.0054	ACAATGTCACAG
MPR.147	PA	40.50319	-75.3174	M	87	16	0.0021	adult	-20.5	6.7	-1.2217	3.1057	GCCATAGTGTGT
MPR.148	PA	40.45099	-75.40128	M	75	13.25	0.0024	adult	-21.6	6.1	-1.2121	3.0214	GGTCCCGAAATT
MPR.150	PA	40.59571	-75.65891	F	75	11.875	0.0021	adult	-23.8	5.9	-1.2531	2.9697	TCTGCGAGTCTG
MPR.151	PA	40.55189	-75.32455	M	78.5	15	0.0024	adult	-23.2	7.1	-1.2217	3.1057	ATGTAGGCTTAG
MPR.152	PA	40.53107	-75.34738	M	81	15	0.0023	adult	-15.4	6.7	-1.2138	3.0054	TGCTTCCAATTC
MPR.153	PA	40.500421	-76.103425	F	81.5	21.25	0.0032	pregnant	-10.7	8.4	-1.3627	3.0875	GCCGAGATAATT
MPR.154	PA	40.461968	-76.08789	F	82	17.25	0.0026	adult	-19.9	6.3	-1.2776	2.9005	TCGAGTATCGAA
MPR.155	PA	40.70747	-75.66738	M	78	14.5	0.0024	adult	-19.9	7.6	-1.3751	3.1802	GCCCTATCTTCT
MPR.156	PA	40.526708	-76.097258	F	96	21	0.0023	adult	-17	6.8	-1.3627	3.0875	AGGTACGCAATT

Table S9. Metada of wild-caught individuals (continued).

ID	Localities	Latitude	Longitude	Sex	Body length	Weight	BMI	Repro. status	Carbon	Nitrogen	ClimatePC1	ClimatePC2	BarcodeSequence
MPR.159	VA	37.74117	-77.57546	F	77	14.75	0.0025	adult	-23.5	7.4	-0.5471	2.6809	GTCCCTATTATC
MPR.161	VA	37.65535	-77.72927	F	80	15.25	0.0024	pregnant	-24.1	8	-0.5269	2.6541	TGGGACATATCC
MPR.162	VA	37.48763	-77.40208	M	86	17.25	0.0023	adult	-19.5	6.5	-0.1458	2.5907	GAACGATCATGT
MPR.163	VA	37.65501	-77.73977	M	78	15.125	0.0025	adult	-22.8	5.9	-0.5269	2.6541	TTCAGACCAGCC
MPR.164	VA	37.67885	-77.5166	M	91	18.25	0.0022	adult	-20.3	6.9	-0.5471	2.6809	ACGCATCGCACT
MPR.165	VA	37.55811	-77.48073	M	66	10.75	0.0025	adult	-19.3	8.2	-0.3668	2.6924	CAGTAGCGATAT
MPR.166	VA	37.76063	-77.48351	M	80	15	0.0023	adult	-19	5.9	-0.4681	2.6642	TCCCTTGTCTCC
MPR.167	VA	37.52756	-77.6492	F	80	13	0.0020	adult	-14.7	9.6	-0.4886	2.7185	ACGAGACTGATT
MPR.168	VA	37.85448	-77.26262	F	82	13.5	0.0020	pregnant	NA	NA	-0.4758	2.7157	GCTGTACGGATT
MPR.169	VA	37.55537	-77.66149	F	69	11	0.0023	adult	-14.1	7.5	-0.4886	2.7185	ATCACCAGGTGT

Table S10. Principal components of climatic variables.

	PC1 (69.2%)	PC2 (22.1%)	PC3 (4.6%)	Prin4 (3.6%)
bio1	0.27111	0.0579	0.11257	0.09989
bio2	0.1859	-0.00256	-0.44368	0.7295
bio3	0.27115	-0.06465	-0.04618	0.12554
bio4	-0.27493	0.02479	-0.03135	-0.02992
bio5	0.25218	0.159	0.19138	0.16051
bio6	0.27079	0.05962	0.14904	0.03134
bio7	-0.27357	-0.00493	-0.12247	0.03852
bio8	0.06238	-0.36347	0.63004	0.23546
bio9	0.25284	0.16393	-0.20573	-0.01738
bio10	0.25949	0.12593	0.18248	0.141
bio11	0.27311	0.03834	0.097	0.07757
bio12	0.25318	0.10785	-0.00326	-0.38185
bio13	0.25923	-0.12326	-0.14616	-0.20514
bio14	-0.01549	0.48149	0.12471	0.04536
bio15	0.21244	-0.2939	-0.18019	-0.14671
bio16	0.25764	-0.13577	-0.10009	-0.24175
bio17	-0.00357	0.46494	0.30991	-0.04143
bio18	0.22948	-0.25794	0.06483	-0.17746
bio19	0.16001	0.37484	-0.22778	-0.16795

Table S11. Metadata of lab reared individuals from the most northern and southern populations.

ID	Breeding ID	Locality	Sex	Weight	Age (days)	Carbon	Nitrogen	BarcodeSequence
FL01F1	F1(DL95x54)	Florida_Lab	F	16.6	146	-19.0	6.7	TTGGTCTCCTCT
FL01F2	F1(DL95x54)	Florida_Lab	F	19	146	-19.0	6.7	CTGCATACTGAG
FL01M1	F1(DL95x54)	Florida_Lab	M	23.6	146	-19.0	6.7	TACCTAGTGAGA
FL01M2	F1(DL95x54)	Florida_Lab	M	21	146	-19.0	6.7	CGTTCTGGTGGT
FL02F1	F1(DL105x71)3	Florida_Lab	F	15.7	194	-19.0	6.7	GTCAATTAGTGG
FL02F2	F1(DL105x71)1	Florida_Lab	F	16.2	264	-19.0	6.7	AGTACGCAGTCT
FL02M1	F1(DL105x71)2	Florida_Lab	M	20.2	242	-19.0	6.7	CAGGGCCTTTGT
FL02M2	F1(DL105x71)	Florida_Lab	M	23.5	194	-19.0	6.7	CGATGAATATCG
FL03F1	F1(DL107x83)	Florida_Lab	F	17.5	194	-19.0	6.7	TTCCCGAAACGA
FL03F2	F1(DL107x83)3	Florida_Lab	F	20.5	265	-19.0	6.7	GAACTTTAGCGC
FL03M1	F1(DL107x83)4	Florida_Lab	M	25	194	-19.0	6.7	AGCAGCTATTGC
FL03M2	F1(DL107x83)6	Florida_Lab	M	27.2	194	-19.0	6.7	CTCGGATAGATC
FL04F1	F1(DL58x59)5	Florida_Lab	F	12.9	169	-19.0	6.7	TACTGAGCCTCG
FL04F2	F1(DL58x59)7	Florida_Lab	F	15.5	169	-19.0	6.7	AGAAGGCCTTAT
FL04M1	F1(DL58x59)	Florida_Lab	M	20	148	-19.0	6.7	TCCTTAGAAGGC
FL04M2	F1(DL58x59)	Florida_Lab	M	20.9	148	-19.0	6.7	GATGGACTTCAA
FL05F1	F1(DL69x77)3	Florida_Lab	F	11.9	126	-19.0	6.7	CATCGCGTTGAC
FL05F2	F1(DL69x77)1	Florida_Lab	F	11.8	126	-19.0	6.7	GCACATAGTCGT
FL05M1	F1(DL69x77)	Florida_Lab	M	19.5	129	-19.0	6.7	GGAATTATCGGT
FL05M2	F1(DL69x77)	Florida_Lab	M	15.8	129	-19.0	6.7	CATCAAGCATAG
FL06F1	F1(DL52x49)	Florida_Lab	F	16.1	177	-19.0	6.7	CCTAGTAAGCTG
FL06F2	F1(DL52x49)	Florida_Lab	F	12.5	100	-19.0	6.7	TTACCGACGAGT
FL06M1	F1(DL52x49)	Florida_Lab	M	19.2	177	-19.0	6.7	GGCAAATACACT
FL06M2	F1(DL52x49)	Florida_Lab	M	14.2	128	-19.0	6.7	GTCATGCTCCAG
FL07F1	F1(DL80x85)5	Florida_Lab	F	13.5	232	-19.0	6.7	TTACCTTACACC
FL07F2	F1(DL80x85)3	Florida_Lab	F	15.6	232	-19.0	6.7	TGACTAATGGCC
FL07M1	F1(DL80x85)	Florida_Lab	M	15.4	148	-19.0	6.7	GCTTAGATGTAG
FL07M2	F1(DL80x85)	Florida_Lab	M	16.9	148	-19.0	6.7	AAGACGTAGCGG
FL08F1	F1(DL67x66)3	Florida_Lab	F	18.5	266	-19.0	6.7	CACGTGACATGT
FL08F2	F1(DL67x66)5	Florida_Lab	F	22	266	-19.0	6.7	CACAGTTGAAGT
FL08M1	F1(DL67x66)	Florida_Lab	M	21.1	205	-19.0	6.7	CTCTCTCACTTG
FL08M2	F1(DL67x66)6	Florida_Lab	M	21.5	205	-19.0	6.7	ATTGCAAGCAAC
FL09F1	F1(DL81x82)1	Florida_Lab	F	15	213	-19.0	6.7	ACCGGAGTAGGA
FL09F2	F1(DL81x82)5	Florida_Lab	F	13	146	-19.0	6.7	TGAGGACTACCT
FL09M1	F1(DL81x82)	Florida_Lab	M	17.5	147	-19.0	6.7	CTAGGATCACTG
FL09M2	F1(DL81x82)2	Florida_Lab	M	20.1	213	-19.0	6.7	GATGACCCAAAT
FL10F1	F1(DL97x73)5	Florida_Lab	F	15.5	200	-19.0	6.7	TGACCGGCTGTT
FL10F2	F1(DL97x73)1	Florida_Lab	F	15	200	-19.0	6.7	GGAGGAGCAATA
FL10M1	F1(DL97x73)	Florida_Lab	M	18.8	179	-19.0	6.7	CAATCGGCTTGC
FL10M2	F1(DL97x73)	Florida_Lab	M	18.4	179	-19.0	6.7	AACACTCGATCG

Table S11. Metadata of lab reared individuals from the most northern and southern populations (continued).

ID	Breeding ID	Locality	Sex	Weight	Age (days)	Carbon	Nitrogen	BarcodeSequence
NY01F1	F1(MJS19x13)5	New_York_Lab	F	17.8	235	-19.1	6.4	AGAGAGACAGGT
NY01F2	F1(MJS19x13)3	New_York_Lab	F	17.9	235	-19.1	6.4	TCGCCAGTGCAT
NY01M1	F1(MJS19x13)8	New_York_Lab	M	20.1	142	-19.1	6.4	TCAACCCGTGAA
NY01M2	F1(MJS19x13)10	New_York_Lab	M	21.5	142	-19.1	6.4	GTTTGAAACACG
NY02F1	F1(MJS66x67)3	New_York_Lab	F	13.5	166	-19.1	6.4	TCTAGCCTGGCA
NY02F2	F1(MJS66x67)1	New_York_Lab	F	16.8	166	-19.1	6.4	AATGCAATGCGT
NY02M1	F1(MJS66x67)	New_York_Lab	M	21	140	-19.1	6.4	GCTCAGGACTCT
NY02M2	F1(MJS66x67)	New_York_Lab	M	24.8	140	-19.1	6.4	CACTTTGGGTGC
NY03F1	F1(MJS34x38)3	New_York_Lab	F	17.9	236	-19.1	6.4	ATCAGAGCCCAT
NY03F2	F1(MJS34x38)1	New_York_Lab	F	22.1	236	-19.1	6.4	TCTGTAGAGCCA
NY03M1	F1(MJS34x38)	New_York_Lab	M	20.1	144	-19.1	6.4	CGAATGAGTCAT
NY03M2	F1(MJS34x38)2	New_York_Lab	M	22	210	-19.1	6.4	CAACGCTAGAAT
NY04F1	F1(MJS44x42)5	New_York_Lab	F	16.4	165	-19.1	6.4	GACAACGAATCT
NY04F2	F1(MJS44x42)9	New_York_Lab	F	17.8	233	-19.1	6.4	TGCGGTTGACTC
NY04M1	F1(MJS44x42)	New_York_Lab	M	22.6	142	-19.1	6.4	CCGACTCTAGGT
NY04M2	F1(MJS44x42)	New_York_Lab	M	23.9	142	-19.1	6.4	ATCCTACGAGCA
NY05F1	F1(MJS49x60)	New_York_Lab	F	17.5	139	-19.1	6.4	CACAGGATTACC
NY05M1	F1(MJS49x60)	New_York_Lab	M	23.4	139	-19.1	6.4	TGAGAAGAAAGG
NY05M2	F1(MJS49x60)	New_York_Lab	M	25.1	139	-19.1	6.4	TCGGATCTGTGA
NY05M3	F1(MJS49x60)	New_York_Lab	M	21.9	139	-19.1	6.4	GCCGGTACTCTA
NY06F1	F1(MJS105x104)3	New_York_Lab	F	15	165	-19.1	6.4	TGTGTTACTCCT
NY06F2	F1(MJS105x104)1	New_York_Lab	F	16.5	165	-19.1	6.4	GGTACCTGCAAT
NY06M1	F1(MJS105x104)	New_York_Lab	M	24.2	141	-19.1	6.4	CGATATCAGTAG
NY06M2	F1(MJS105x104)	New_York_Lab	M	25.2	141	-19.1	6.4	CATAAGGGAGGC
NY07F1	F1(MJS9x11)5	New_York_Lab	F	20.1	154	-19.1	6.4	TAACCCGATAGA
NY07F2	F1(MJS9x11)7	New_York_Lab	F	20	154	-19.1	6.4	GTGTGCTAACGT
NY07M1	F1(MJS9x11)	New_York_Lab	M	20.9	133	-19.1	6.4	TCGCCTATAAGG
NY07M2	F1(MJS9x11)	New_York_Lab	M	24.8	180	-19.1	6.4	AGTGGCACTATC
NY08F1	F1(MJS83x80)	New_York_Lab	F	13.8	203	-19.1	6.4	ATTGCTGGTCGA
NY08F2	F1(MJS83x80)1	New_York_Lab	F	15	226	-19.1	6.4	AAGAAGCCGGAC
NY08M1	F1(MJS83x80)	New_York_Lab	M	17	130	-19.1	6.4	CTTGCGGCAATC
NY08M2	F1(MJS83x80)4	New_York_Lab	M	16.2	203	-19.1	6.4	TGAGGTTTGATG
NY09F1	F1(MJS82x81)1	New_York_Lab	F	21.1	226	-19.1	6.4	TCCGTCATGGGT
NY09F2	F1(MJS82x81)7	New_York_Lab	F	23.2	169	-19.1	6.4	AGATCTATGCAG
NY09M1	F1(MJS82x81)4	New_York_Lab	M	24.5	204	-19.1	6.4	ACGGGATACAGG
NY09M2	F1(MJS82x81)	New_York_Lab	M	27	147	-19.1	6.4	AAGAGTCTCTAG
NY10F1	F1(MJS92x91)5	New_York_Lab	F	17	226	-19.1	6.4	GCGAGTTCCTGT
NY10F2	F1(MJS92x91)	New_York_Lab	F	20.9	168	-19.1	6.4	TTCCGAATCGGC
NY10M1	F1(MJS92x91)2	New_York_Lab	M	24.2	204	-19.1	6.4	GCACAAGGCAAG
NY10M2	F1(MJS92x91)	New_York_Lab	M	25.5	144	-19.1	6.4	CGGCAAACACTT

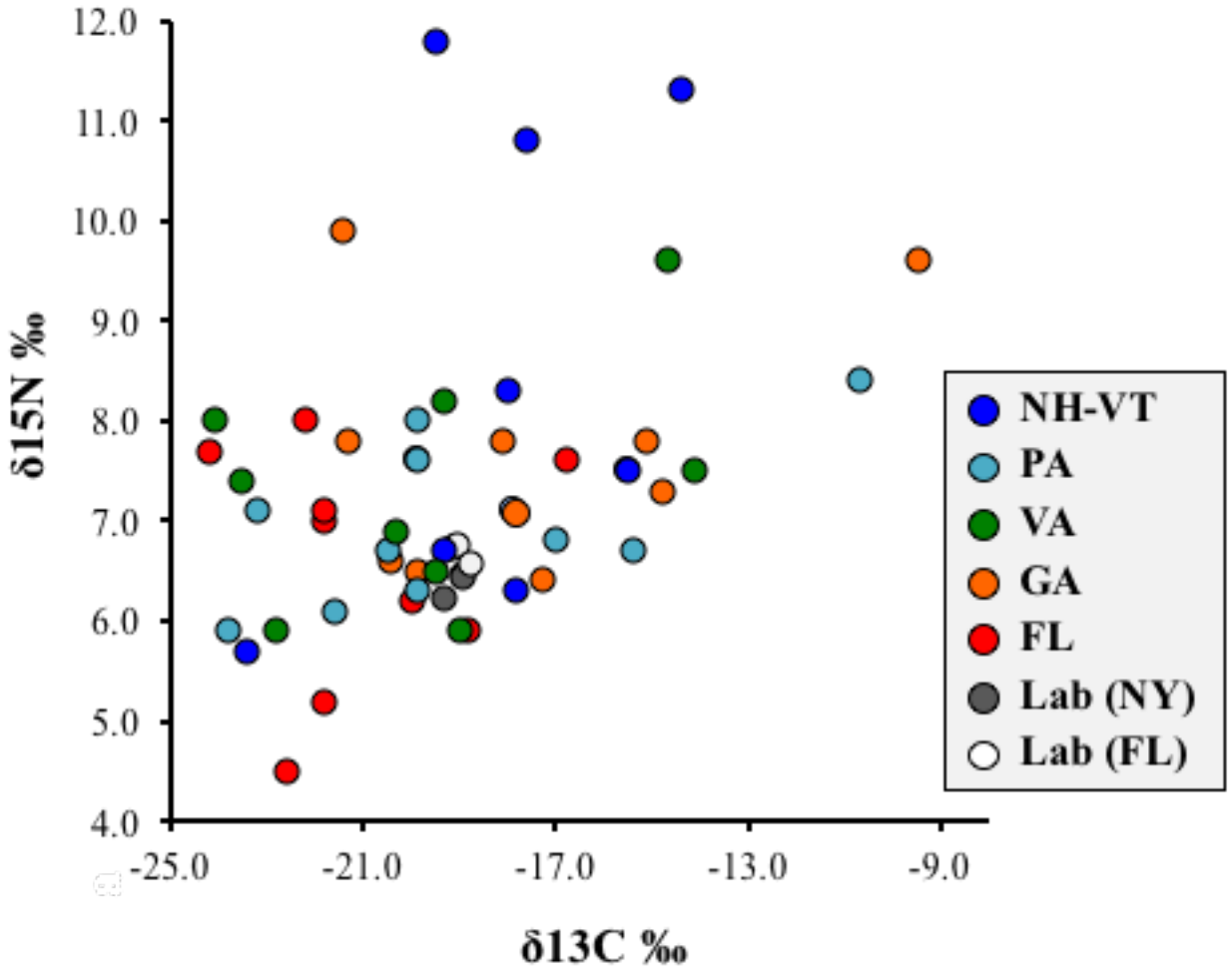


Figure S1. Dietary difference among wild and lab populations. No significant differences were observed among the populations (Kruskal-Wallis tests: δ13C ‰, p-value = 0.27 and δ15N ‰, p-value = 0.24).

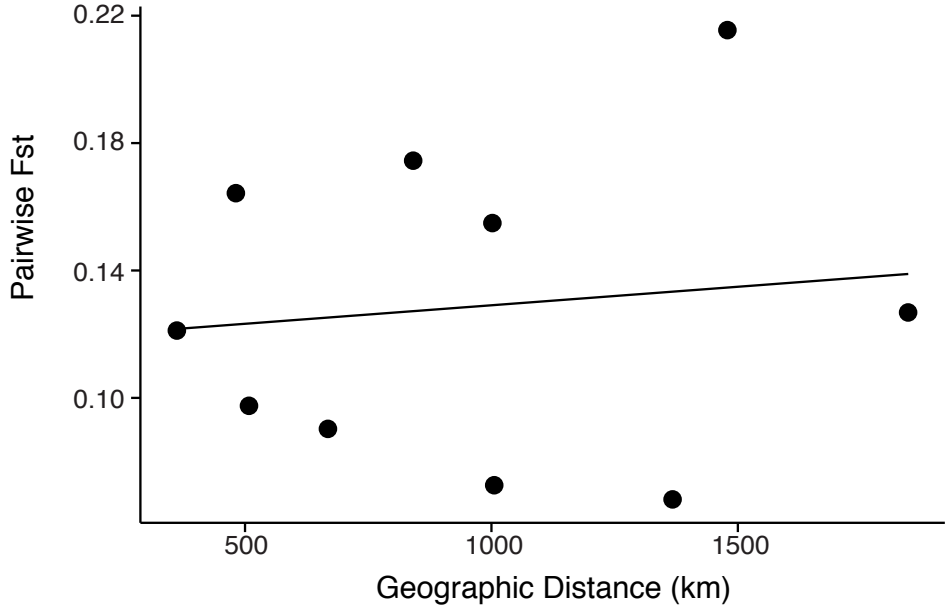
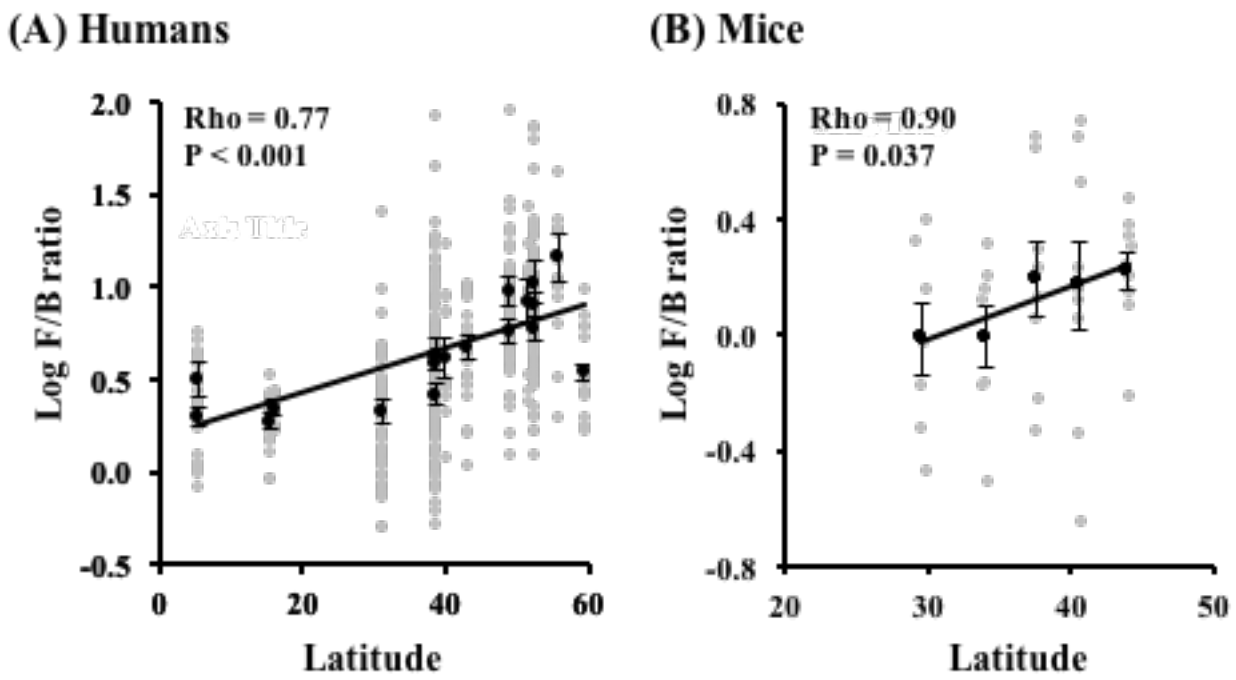


Figure S2. There is no evidence of a significant association between genetic distance and geographic distance (Mantel test, Exome: Z= 1225.03, r = 0.12, *P* = 0.37). Figure adopted from Phifer-Rixey *in review*.



FigureS3. Correlations between obesity-associated composition (Log Firmicutes/Bacteroidetes ratio) and latitudes. Black dots are population average and gray dots are individual data points. (A) Figure modified from Suzuki and Worobey 2014. Significant positive correlation for population average (Rho = 0.77, P < 0.0001) and individual data points (Rho = 0.51, P < 0.0001) in healthy adult human populations. (B) Significant positive correlation for population average (Rho = 0.90, P = 0.037), but not for individual data points (Rho = 0.26, P = 0.11) in adult mouse populations. There are also weak trends of positive correlations between Log F/B ratio and Body weight (rho = 0.29, p=0.0693) and BMI (rho = 0.282, p = 0.0779) in wild mice. Error bars are SE.

	P > 0.1 P < 0.1 P < 0.05 P < 0.01	Spearman's Rho **Wild**		
		Latitude	BMI*	Body weight*
Alpha-diversity				
Phylogenetic diversity		0.401	0.254	0.410
Phylum Firmicutes		0.134	0.210	0.081
Genus *Clostridiales_unc*		0.078	0.145	0.101
Genus *Lachnospiracea_unc*		0.128	0.160	0.138
Genus *Lactobacillus*		0.047	0.155	0.031
Genus *Oscillospira*		-0.069	0.290	0.033
Genus *Ruminococcaceae_unc*		0.081	0.086	0.116
Genus *Ruminococcus*		-0.115	-0.144	-0.024
Phylum Bacteroidetes		-0.336	-0.160	-0.155
Genus *Bacteroides*		-0.061	-0.169	-0.128
Genus *Bacteroidales_unc*		0.196	0.225	0.144
Genus *Odoribacter*		0.097	0.274	0.412
Genus *Parabacteroides*		-0.203	-0.397	-0.180
Genus *Prevotella*		0.140	-0.390	-0.200
Genus *Rikenellaceae_unc*		-0.036	0.256	0.228
Genus *S24_7_unc*		-0.429	-0.099	-0.004
Phylum Proteobacteria		0.443	-0.050	0.219
Genus *Desulfovibrionaceae_unc*		0.372	0.073	0.249
Genus *Helicobacteraceae_unc*		0.198	-0.055	0.151
Genus *Helicobacter*		0.374	-0.155	-0.207
Phylum Deferribacteres		0.266	0.149	0.033
Genus *Mucispirillum*		0.266	0.149	0.033

Figure S4. Correlations between microbial measurements and metadata (i.e. latitude, BMI, and body weight). * BMI and body weight are controlled for latitude by using the residuals of covariate regression. Spearman's rho values are shown and the p-values are indicated by color. The genera are chosen based on average relative abundance of >1% in the wild samples. Pregnant females and juveniles are excluded and only adults (n=40) are included in the analyses.

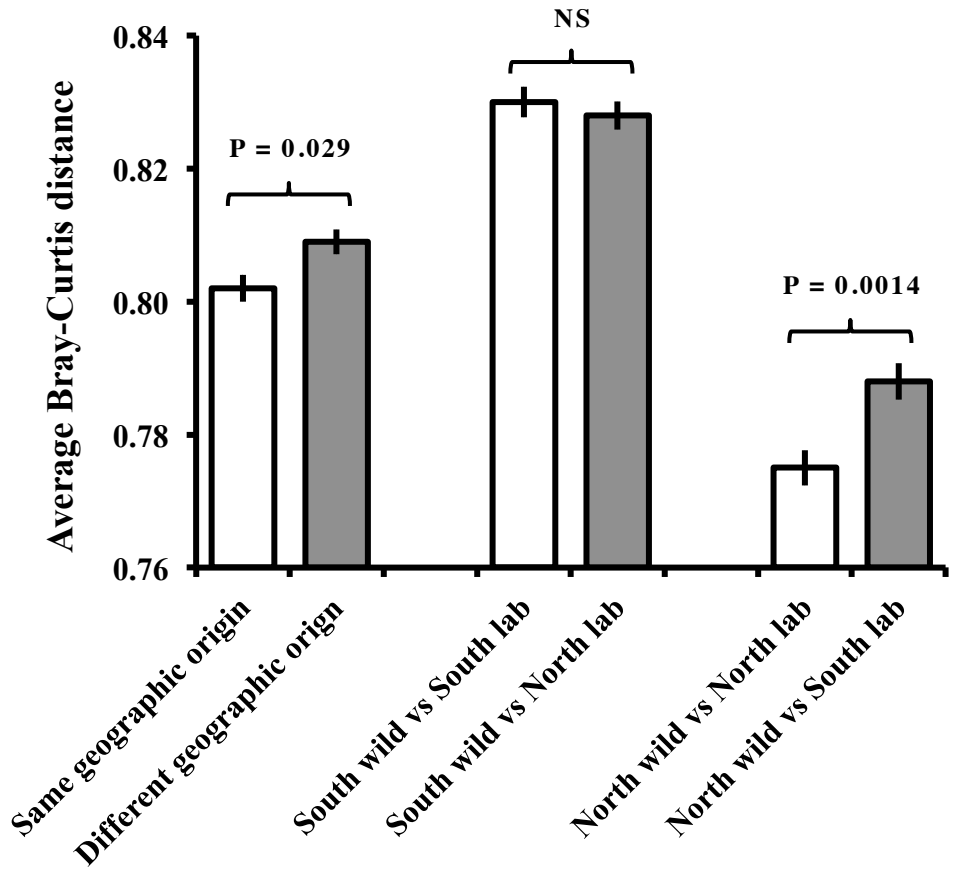


Figure S5. The microbial community composition is more similar between wild and lab reared animals that share the same geographic origin compared to those that do not share the geographic origin ($p = 0.029$). The southern wild population was equally similar to the two lab reared populations ($p = 0.44$), but the northern wild population was significantly more similar to northern lab population compared to southern lab population ($P=0.0014$). P-values are Wilcoxon permutation test based on 9999 Monte-Carlo resampling. Error bars are SE.

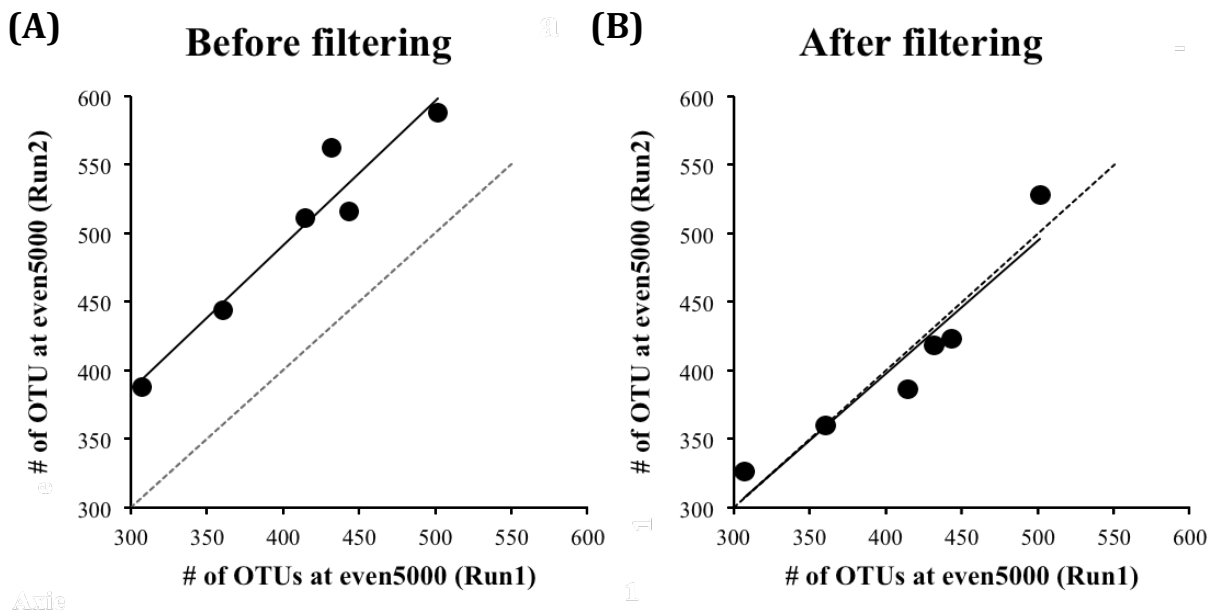


Figure S6. Correlations between OTU counts on the same six samples between two lanes of Illumina MiSeq before and after filtering. Dotted lines show the expected regression between the two samples ran on different lanes of Illumina sequencing. (A) Run2 shows greater number of OTU counts compared to run1 despite rarefying to an even depth of 5000 reads. (B) Rare OTUs with relative abundance less than 8.0×10^{-6} were removed from the run2 OTU table to account for the lane bias.

Chapter 5

The role of atmospheric oxygen on the assembly and function of the gut microbiota in high-altitude mammals

Anticipated co-authorship: Felipe Martins[1], Michael Nachman[2]

[1] *Research School of Biology, The Australian National University, Canberra, Australia.*
[2] *Department of Integrative Biology and Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, CA, 94720, USA.*

Abstract

The maintenance of oxygen homeostasis in the gut is critical for the maintenance of a healthy gut microbiota. Variation in the partial pressure of oxygen in the gut predicts different abundances of anaerobic and aerobic bacteria, but how the atmospheric oxygen concentration affects the gut microbiota in natural populations remains unexplored. High altitude environments provide an opportunity to study the potential effects of atmospheric oxygen on the composition and function of the gut microbiota. Here, we tested the effects of altitude on the gut microbiota by characterizing the cecal microbial communities of wild house mice from two independent transects in Ecuador and Bolivia, where the elevation ranges from sea level to 3906m. First, we found that differences in altitude were associated with differences in the gut microbial community after controlling for the effects of body size, diet, reproductive status, and population. Second, obligate anaerobes tended to show positive correlations with altitude while all other microbes tended to show negative correlations with altitude. These patterns were seen independently in both transects, consistent with the expected effects of atmospheric oxygen on the gut microbiota. *Prevotella* was the genus that was most enriched in high-altitude environments in both transects, consistent with observations in high-altitude populations of pikas, ruminants, and humans, and also consistent with laboratory mice exposed to hypoxic conditions. Lastly, the renin-angiotensin system, a recently proposed microbiota-mediated pathway of blood pressure regulation, was the top predicted metagenomic pathway enriched in high altitudes in both transects. These results suggest that high altitude environments may affect the compositional and functional variation of the gut microbiota in wild mammals.

5.1. Introduction

The gut microbiota can affect the health (Knight et al., 2017) and fitness (Suzuki, 2017) of animals. The maintenance of oxygen homeostasis in the gut is critical in maintaining a healthy gut microbiota (Zeitouni, Chotikatum, von Köckritz-Blickwede, & Naim, 2016). A healthy adult mammalian gut microbiota is generally dominated by anaerobes and few aerobes. Alteration of the gut microbial community can lead to

expansions of pathogenic aerobic bacteria (Byndloss et al., 2017; Rivera-Chávez et al., 2016). Thus, identifying the factors that affect the composition and function of the gut microbiota has broad implications.

The oxygen concentration in the gut is one of the key factors that affect the assembly of the gut microbial community. The first colonizers of the infant gut are often aerobes that consume oxygen, and later colonizers tend to be obligate anaerobes (Matamoros, Gras-Leguen, Le Vacon, Potel, & De La Cochetiere, 2013; Palmer, Bik, DiGiulio, Relman, & Brown, 2007). Spatial variation of gut microbial communities also exists within an individual's gastrointestinal (GI) tract. For example, consistent with the decrease in oxygen levels from the mouth to anus (He et al., 1999), aerobes tend to dominate the upper GI tract and anaerobes tend to dominate the lower GI tract in mice (Gu et al., 2013; Suzuki & Nachman, 2016) and humans (Hayashi, Takahashi, Nishi, Sakamoto, & Benno, 2005). Within the lower GI tract, anaerobes dominate the luminal content and aerobes are enriched in the oxygen-rich mucus layer where the oxygen diffuses from epithelial cells as seen in mice, macaques, and humans (Espey 2013; Albenberg et al. 2014; Yasuda et al. 2015). Hypoxic exposure in laboratory mice can induce changes in the gut microbial composition, including an increase in obligate anaerobes suggesting that an oxygen deficit can provide a selective advantage to anaerobes over aerobes (Moreno-Indias et al., 2015). Although other factors are well-known to influence the gut microbial composition in natural populations of mammals including diet (Wang et al., 2014; Wu et al., 2011), host genetics (Goodrich, Davenport, Waters, Clark, & Ley, 2016), body size (Ley, Turnbaugh, Klein, & Gordon, 2006; Nishida & Ochman, 2017), and reproductive status (Nuriel-Ohayon, Neuman, & Koren, 2016), how atmospheric oxygen levels affect gut microbial composition has been less explored outside of laboratory settings.

High altitude environments provide an opportunity to study the effects of atmospheric oxygen on the compositional and functional variation of the gut microbiota. The reduced partial pressure of oxygen at higher elevations causes a variety of physiological issues related to hypoxic stress in animals (Grindlay & Regensteiner, 1983; Storz, 2007). A few studies have characterized the gut microbiota from mammalian populations living at different altitudes, including work on pikas (H. Li, Li, Beasley, et al., 2016; H. Li, Li, Yao, et al., 2016), ruminants (Zhang et al., 2016), macaques (Sun et al., 2016), and humans (Lan et al., 2017; K. Li et al., 2016; L. Li & Zhao, 2015). In all of these studies, differences in altitude were associated with differences in the gut microbial composition. However, the observed patterns could also be explained by differences in diet, population structure, or culture, as these variables co-varied with altitude. Disentangling the effects of atmospheric oxygen and other co-variables on the gut microbiota remains a challenge.

Two major beneficial functions of the gut microbiota in high altitude environments have been proposed; the gut microbiota-mediated energy harvest (Lan et al., 2017; K. Li et al., 2016; L. Li & Zhao, 2015; Zhang et al., 2016; Zhao et al., 2018) and blood pressure regulation (Lan et al., 2017; L. Li & Zhao, 2015). Many anaerobic gut bacteria produce short-chain fatty-acids (SCFAs) as end products of polysaccharide fermentation (Topping & Clifton, 2001). SCFAs are a major energy source for epithelial cells and provide about 10% of daily calories in humans (Bergman, 1990). At high altitudes, a greater abundance of SCFA-producing obligate anaerobes was reported (Lan

et al., 2017; K. Li et al., 2016; L. Li & Zhao, 2015; Zhang et al., 2016; Zhao et al., 2018). SCFAs are not only an important energy source, but they also act as signaling molecules by traveling through the bloodstream and binding to SCFA-receptors in various tissues (Samuel et al., 2008). Pluznick et al. (2014) proposed opposing roles of two SCFA-receptors that decrease and increase blood pressure. For example, olfactory receptor 78 (*Olfr78*) in the kidney can increase blood pressure through SCFA-mediated renin release, an enzyme that plays a central role in regulating blood pressure through the renin-angiotensin system (J L Pluznick et al., 2009; Jennifer L Pluznick et al., 2013). These discoveries led some researchers to speculate that the gut microbiota may be involved in the regulation of blood pressure in high-altitude populations of humans (Lan et al., 2017; L. Li & Zhao, 2015).

House mice provide an opportunity to study the effects of a high-altitude environment on the composition and function of the gut microbiota. First, house mice (*Mus musculus*) are predominantly a lowland species, but they have successfully colonized high-altitude environments in the last few hundred years with human settlers including at elevations over 4000m in Peru (Harland, 1958) and Bolivia (Storz et al., 2007). The ability to study multiple altitudinal transects from sea level to over 4000m elevation within a single species provides an opportunity to look for parallel patterns in the gut microbiota in relation to altitude. Second, although the extent to which house mice have adapted to hypoxic environments remains unclear (Storz et al., 2007), there is evidence that lab mice kept in high-altitudes for 30 generations show differences in their physiology compared to mice from low-altitudes (A. Jochmans-Lemoine et al., 2015; Alexandra Jochmans-Lemoine, Shahare, Soliz, & Joseph, 2016). Lastly, house mice are good mammalian models (Phifer-Rixey & Nachman, 2015). Genomic and physiological research related to hypoxia and the microbiota in laboratory mice provides useful information for interpreting patterns observed in natural populations of wild mice.
Here we characterized the cecal microbial community of natural populations of house mice across two altitudinal transects in South America to test the effects of atmospheric oxygen on the gut microbiota. First, we tested whether altitude correlates with overall differences in the gut microbiota independent of covariates. Second, we tested whether obligate anaerobes show positive correlations with altitude and whether aerobes show negative correlations with altitude. Third, we looked for parallel patterns in the gut microbiota across high-altitude mammals and laboratory mice under hypoxic exposure. Lastly, we identified predicted metagenomic pathways that correlate with altitude in both transects to generate hypotheses on the gut microbial functions in high-altitude environments. Overall, the results were consistent with the idea that atmospheric oxygen alters gut microbial composition. We also discuss potential beneficial effects of the gut microbiota in high-altitude environments.

5.2. Material and Methods

5.2.1. Sampling

Ten populations of house mice (*Mus musculus*) were collected across two altitudinal transects in South America including a total of 92 individuals (Fig.1). The Ecuador transect included five populations; Portoviejo (n = 11, mean elev. = 32m), Santo

Domingo (n = 10, mean elev. = 416m), Nanegalito (n= 9, mean elev. = 1643m), Tumbaco (n = 11, mean elev. = 2598m), and Latacunga (n = 8, mean elev. = 2918m) (Fig.1A). The Bolivia-Brazil transect also included five populations; Porto Velho (n = 9, mean elev. = 86m), Santa Cruz (n = 4, mean elev. = 320m), Cochabamba (n = 12, mean elev. 2615m), La Paz (n = 10, mean elev. 3435m), and Lake Titikaka (n = 8, mean elev. 3846m) (Fig. 1B). Individuals were collected using Sherman live traps and each individual was separated at least 500m to avoid collecting close relatives except four sites in Ecuador transect where two to three individuals were collected from the same site (Table S1). All of the Ecuador samples were collected in Nov. – Dec. 2012. All of the Bolivia-Brazil samples were collected in Aug. – Sep. 2014 except the Porto Velho population which were collected in Sep. 2013.

Cecal samples and external measurements including body weight and body mass index (i.e. BMI = (body weight / body length)^2) were collected within 24 hours after capture. The cecal samples were stored in RNAlater solution at 4°C overnight, transferred to liquid nitrogen after 8-12 hours except the samples from Porto Velho where the cecal samples were directly stored in liquid nitrogen. All the samples were stored in deep freezer (-80°C) until sequencing. While different sample preservation methods are known to affect the microbial community composition (Choo, Leong, & Rogers, 2015), we decided to include the Porto Velho population to increase the power of the analyses and accounted for the effect of sample preservation statistically (see below). Carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) stable isotope ratios were analyzed from mouse hair to estimate diet following the protocol of Suzuki and Nachman (2016). Detailed information of individual measurements is summarized in Table S1. All procedures involving animals were reviewed and approved by the Institutional Animal Care and Use Committee at the University of Arizona (protocol 07-004) and the University of California Berkeley (protocol R361-0514). Animal collection permits were issued by the local governments where necessary. Museum specimens (skins and skulls) were prepared and have been deposited in the mammal collection of the Museum of Vertebrate Zoology at the University of California, Berkeley and the data were uploaded to a public database ARCTOS.

5.2.2. DNA extraction and 16S rRNA gene sequencing

The distal portion of cecal samples including both luminal and mucosal samples were weighted equally (~200mg) in sterile condition. We used a protocol of DNA extraction described in Suzuki and Nachman (2016). Briefly, we added mechanical disruption step using sterile zirconia/silica beads (0.1mm, Research Products International Corp.) before step 4 (vortex and centrifugation) in the protocol from the QIAamp DNA stool Minikit (Qiagen). The samples were stored at -20°C before sequencing. The V4 region of the 16S rRNA gene was amplified, multiplexed, and sequenced for 16S amplicon sequencing on 150bp pair-end Illumina MiSeq at the Next Generation Sequencing Core Facility at Argonne National Laboratory.
The PCR primers (515F and 806R) and the barcodes are described in (Caporaso *et al.* 2012).

5.2.3. Data analyses

All of the 16S rRNA data was analyzed in QIIME version 1.9.0 (Caporaso *et al.* 2010). The forward reads were demultiplexed and quality-filtered using default parameters using *split_libraries_fastq.py*. Subsampled open-reference OTU picking approach (*pick_open_reference_otus.py*) was used with default parameters. UCLUST was used to generate 97% OTUs (Edgar 2010) and taxa were assigned based on SILVA database (release 128) (Quast *et al.* 2013). Singletons were removed. A phylogenetic tree was created using FastTree (Price, Dehal and Arkin 2009). The OTU table was rarefied to an even depth of 5,000 reads to maximize the sample size.

Bray-Curtis dissimilarity was calculated and PCoA plots were generated including all individuals using *jackknife_beta_diversity.py* with default parameters. To identify variables that significantly affect the Bray-Curtis dissimilarity, we used ADONIS with 9999 permutations. We calculated correlations between altitude and other variables in the metadata by using Spearman's rho correlation. Population difference in diet stable isotope measurements were tested using ANOVA. P-values of ADONIS, Spearman's correlation, and ANOVA were corrected for multiple testing using Bonferroni correction. To test whether the correlation between altitude and Bray-Curtis dissimilarity is independent of other covariates, we conducted model comparisons using linear mixed-effects model with "lmer" function in the package "lme4" in R (version 3.4.3). All of the variables were normalized using standard deviation. The first three principle components of Bray-Curtis dissimilarity, PC1 (8.4%), PC2 (6.2%), and PC3 (5.3%) were used as response variables. The full models included five fixed effects (i.e. altitude, body weight, BMI, carbon and nitrogen stable isotope diet measures) and three random effects (i.e. population, reproductive status, and sample storage methods). The full models were compared to models without altitude using Akaike information criterion with sample size correction (AICc) with the "AICc" function in the package "AICcmodavg". We also used Likelihood ratio test to compare between full models and models without altitude with "lrtest" function in the package "lmtest".

The overall alpha-diversity measurements (i.e. OTU counts, phylogenetic diversity, and simpson's index) were calculated based on the rarefied OTU table using *alpha_diversity.py*. In addition, to identify bacterial phyla that show increase in alpha-diversity in high-altitude environments, we calculated the phylogenetic diversity within each of the three dominant phyla (i.e. Firmicutes, Bacteroidetes, and Proteobacteria). Correlations between altitude and alpha-diversity measurements were based on Spearman's rho correlation. To account for covariates, we conducted model comparisons using linear mixed-effects models with the same variables mentioned above except the response variables being the phylogenetic diversity.

The relative abundances of bacterial phyla and genera were calculated based on the rarefied OTU table using *summarize_taxa.py*. We focused on bacterial taxa that have relative abundance of 0.1% or greater and tested the correlations between altitude and relative abundance of bacterial taxa within each transect independently using Spearman's rho correlation. When the slopes between altitude and relative abundances of bacterial taxa were in the same direction in both transects, Fisher's method of combining p-values corrected by Bonferroni correction were used to rank the bacterial taxa that correlated with altitude. Oxygen requirements of bacterial genera (i.e. obligate anaerobes, facultative anaerobes, aerotolerant anaerobes, microaerophiles, and obligate aerobes)

were assigned based on Bergey's Manual of Systematics of Archaea and Bacteria (Whitman *et al.* 2015) and recent literature (Robertson *et al.* 2005; Hardham *et al.* 2008; Reunanen *et al.* 2015; Ouwerkerk *et al.* 2016). The oxygen requirements of bacterial genera were divided into two groups for statistical purposes, "obligate anaerobes" and "all other oxygen requirement types" as in (Albenberg *et al.* 2014). When the genera were unclassified, we used the oxygen requirements of the family. There were three instances where the family of the unclassified genera included both obligate anaerobes and all other oxygen requirement types. In those cases, we searched for all the recognized genera within the family and assigned oxygen requirement based on a majority rule (i.e. two out of the three genera showed the same oxygen requirements).

Lastly, to generate hypotheses of the function of gut microbiota in high-altitudes, we identified predicted metagenomic functions that are significantly correlated with altitude using PICRUSt (Langille *et al.* 2013). Rarefied OTU table with OTUs that are present in the reference database (Greengenes 13_5) were used as an input of PICRUSt. Kyoto Encyclopedia of Genes and Genomes (KEGG) ortholog abundances were assigned and collapsed by KEGG pathway hierarchy level 3. Correlations between relative abundances of KEGG pathway categories and altitude were calculated using Spearman's rho correlation in both transects. Similar to the clinal tests of bacterial genera, when the slopes between altitude and relative abundances of KEGG pathway categories were in the same direction in both transects, Fisher's method of combining p-values with Bonferroni correction were used to rank the bacterial taxa that correlated with altitude.

5.3. Results

5.3.1. Differences in altitude are associated with differences in the gut microbial composition.

To understand the role of altitude affecting the variation of the gut microbiota, we characterized the microbial composition of cecal samples from 10 populations of house mice across two altitudinal gradients, Ecuador transect (Fig.1A) and Bolivia-Brazil transect (Fig.1B). First, we identified factors that significantly explain the differences in the gut microbial community using Bray-Curtis dissimilarity without controlling for covariates. Overall, we found Bray-Curtis dissimilarity was significantly explained by differences in altitude and populations using all samples after correcting for multiple testing (Fig.1C&D, Table S2). Body mass index (BMI) showed significant effects on Bray-Curtis dissimilarity in the Bolivia-Brazil transect, but not in the Ecuador transect (Table S2). Carbon and nitrogen stable isotope diet measurements, body weight, pregnancy, and sex showed no significant associations with Bray-Curtis dissimilarity (Table S2).

The observed association between altitude and Bray-Curtis dissimilarity can be explained by many covariates because the metadata measured in this study were correlated with altitude (Table S3). For example, animals living in higher altitude tend to be heavier ($rho = 0.34$, $p < 0.0001$) and have greater BMI ($rho = 0.33$, $p = 0.001$) consistent with a trend reported in an altitudinal transect of house mice in Peru (Harland 1958). Diet measurements based on carbon and nitrogen stable isotopes significantly

differ between transects and among populations (Fig. S1). There is a trend where carbon and nitrogen stable isotope measurements tend to correlate positively in Ecuador transect and negatively in Bolivia-Brazil transect, although the correlations are not significant (Fig. S1, Table S3).

To test whether altitude is significantly associated with Bray-Curtis dissimilarity independent of other metadata we measured, we conducted model comparisons using linear mix-effects models. We used Bray-Curtis dissimilarity PC1, PC2, and PC3 as response variables. Altitude, body weight, BMI, carbon and nitrogen stable isotope measurements were used as fixed effects and population, pregnancy, and sample storage method were used as random effects. We found that the models without altitude were significantly worse than the full models based on AICc for Bray-Curtis dissimilarity PC1, PC2, and PC3 (Fig. S2). Especially with models with Bray-Curtis dissimilarity PC2 and PC3 as response variables, removing altitude significantly affected the full model based on likelihood ratio tests ($p < 0.05$). The results show that altitude has a significant effect on Bray-Curtis dissimilarity independent of other covariates including diet, body size, and populations.

We next asked whether altitude correlates with alpha-diversity of the gut microbiota (Fig. 1E-H). Overall, alpha-diversity measurement using phylogenetic diversity showed a weak trend of positive correlation with altitude (Fig. 1E, rho = 0.19, p = 0.07). Although the trend of greater species richness at higher altitude was true using other alpha-diversity measurements, none of the correlations were significant after multiple testing (Table S4). To test whether alpha-diversity of individual phyla also show a trend of positive correlation with altitude, phylogenetic diversity was calculated within each of the three dominant phyla (Fig. 1E-H). Only the phylogenetic diversity of Bacteroidetes showed a significant positive correlation with altitude (Fig. 1G). This correlation was mostly driven by Bolivia-Brazil transect (rho = 0.45, p = 0.003) and not by the Ecuador transect (rho = 0.18, p = 0.22). Using a similar linear mixed-effects model comparison mentioned above, the correlation between phylogenetic diversity of Bacteroidetes and altitude remained significant after controlling for other covariates (likelihood ratio test p = 0.015).

Together, the results suggest that differences in altitude were associated with differences in alpha- and beta-diversity of the gut microbial communities in natural populations of house mice.

5.3.2. Oxygen requirements of the bacteria predict the correlations between bacterial genera and altitude.

To identify bacterial taxa that show repeated clinal patterns across altitudinal gradients, we focused on 10 phyla and 38 genera that had average relative abundances of 0.1% or greater across all samples. On average, Firmicutes, Bacteroidetes, and Proteobacteria composed more than 94% of the total gut community (Fig. S3), a typical microbial composition of wild house mice (Linnenbrink *et al.* 2013; Wang *et al.* 2014; Weldon *et al.* 2015; Suzuki and Nachman 2016). None of the bacterial phyla showed significant correlations with altitude after correcting for multiple testing (Table S5). Cyanobacteria showed a trend of positive correlation in the Ecuador transect (rho = 0.29,

p = 0.04) and in the Bolivia-Brazil transect (rho = 0.31, p = 0.04) independently (Table S5).

To test the hypothesis that obligate anaerobes have advantage over all other oxygen requirement types in hypoxic environment, we identified bacterial genera that show repeated patterns of clinal variation with altitude in both transects and assigned oxygen requirements based on recent literature and Bergey's Manual of Systematics of Archaea and Bacteria (Fig.2, Table S6). Among the 38 bacterial genera, 23 of them showed slopes in the same direction with altitude in both transects based on Spearman's rho correlation (Table S6). We were able to assign oxygen requirements to 17 out of the 23 bacterial genera, and 15 of them showed correlations with altitude in the expected directions (Fig.2, sign test, p = 0.002). Even when we excluded the three unclassified genera that we cannot confidently call the oxygen requirements (see methods and Table S6), the pattern remained significant (sign test = 0.01). If we only limit to taxa that show Fisher's combined p-value of less than 0.1, the correlations were all in the expected directions without exceptions (Fig.2). For example, *Akkermansia*, an aerotolerant mucin-degrader that colonizes the mucus layer (Reunanen *et al.* 2015; Ouwerkerk *et al.* 2016), tend to be more abundant in lower altitudes compared higher altitudes. The results are consistent with the hypothesis that reduced atmospheric oxygen at high altitude environment provides advantage to obligate anaerobes over other oxygen requirement types, and vice versa.

5.3.3. High altitude selects for greater relative abundance of *Prevotella* across multiple species of mammals

The top genus that significantly associated with high-altitude environment independently in both transects was the genus *Prevotella* (Fig.2, Table S6). In both Ecuador and Bolivia-Brazil transects, the relative abundance of *Prevotella* increased from 2500m in elevation (Fig. 3A&B). Surprisingly, Li et al. (2016) observed a similar pattern in an altitudinal gradient of pika populations ranging from 1000m to 4331m in elevation (Fig. 3C). In pika cecal samples, the top correlation between altitude and dominant bacterial genera was also *Prevotella* and this was independent of soil- and plant-associated microbial communities (Li *et al.* 2016c). Furthermore, the high-altitude adapted Yaks and Tibetan sheep show greater relative abundance of *Prevotella* compared to their lowland close relatives (Fig.3D and E). A similar association between *Prevotella* and high-altitude has been observed in humans in some studies (Li *et al.* 2016d; Lan *et al.* 2017), but not in others (Li and Zhao 2015). For example, Li, Dan, et al. (2016) found that Tibetan populations living in high-altitudes were dominated by *Prevotella* compared to Han populations living in low-altitudes (Fig. 3D). When bacterial taxa were compared between Tibetan and Han populations living in the same altitude, *Prevotella* was the most enriched taxa in Tibetan populations (Li *et al.* 2016d). Together, the association between *Prevotella* and high altitude environment has been reported at least in house mice, pikas, cattle, sheep, and humans.

The observation between *Prevotella* and high altitude environment can be explained by diet, climate, atmospheric oxygen, or any other variables that correlate with altitude. At least in house mice, diet is unlikely driving this pattern because the correlations between altitude and diet stable isotope measurements tend to be in the opposite directions among

the two transects (Table S3). In contrast, we cannot exclude the possibility of climate driving the bacteria-altitude associations observed in the five mammalian systems because climate and altitude are generally tightly correlated (Grindlay and Regensteiner 1983). However, a study using laboratory mouse model of sleep apnea provides a strong evidence that reduced atmospheric oxygen alone can cause an increase in *Prevotella* (Moreno-Indias *et al.* 2015) (Fig.3E). Moreno-Indias et al. (2015) compared the gut microbiota of laboratory mice in a common environment between intermittent hypoxia and normoxia treatments. They found 6 out of 23 genera significantly differed in their relative abundances between the two treatments including *Paraprevotella* and *Prevotella* that were enriched in intermittent hypoxia treatment relative to controls (Moreno-Indias *et al.* 2015). The results provide experimental evidence that the increase in *Prevotella* at high altitudes can be driven by lower atmospheric oxygen levels.

5.3.4. Predicted metagenome functions suggest gut microbiota-mediated regulations of blood pressure

To generate hypotheses on the functional role of the gut microbiota in host health and fitness in high altitude environment, we looked for predicted KEGG pathways that correlated with altitude in both transects using PICRUSt (Langille *et al.* 2013). Among the 273 KEGG pathways (level 3) that were identified in both transects, 183 of them showed slopes in the same direction with altitude in both transects. Nineteen out of the 183 KEGG pathways showed significant Fisher's combined p-values without correcting for multiple testing (raw p-value < 0.05) (Fig. 4). The top two KEGG pathways that showed positive correlations with altitude were "Renin-angiotensin system" (Ecuador; rho = 0.34, p = 0.02. Bolivia-Brazil; rho = 0.51, p = 0.0005. Fisher's combined p-value = 0.0001) and "hypertrophic cardiomyopathy" (Ecuador; rho = 0.20, p = 0.17. Bolivia-Brazil; rho = 0.53, p = 0.0003, Fisher's combined p-value = 0.0006) (Fig. 4). In contrast, the top KEGG pathway that negatively correlated with altitude was "Glycosphingolipid biosynthesis - lacto and neolacto series" (Ecuador; rho = -0.40, p = 0.004. Bolivia-Brazil; rho = -0.41, p = 0.006. Fisher's combined p-value = 0.0003) (Fig. 4).

After correcting for multiple testing (Bonferroni corrected alpha = 0.05/183 = 0.0003), the only correlation that remained significant was the "renin-angiotensin system" that plays a major role in blood pressure homeostasis (Sparks *et al.* 2014). The significant positive correlation between altitude and "renin-angiotensin system" was driven by three predicted bacterial homologs that are present in diverse organisms including vertebrates and bacteria; angiotensin I converting enzyme (K01283), prolyl endopeptidase (K01322), and thimet oligopeptidase 1 (K01392) (Fig. S5). Computational and biochemical studies suggest that the homologous bacterial enzymes mentioned above may have similar functions in vertebrates (Rivière *et al.* 2007; Sugihara *et al.* 2007; Kaushik and Sowdhamini 2014). All three predicted KEGG orthologs showed positive correlations with altitude in both transects and two of them showed significant Fisher's combined p-value (< 0.05) based on Spearman's rho correlations (Fig. S4). Especially, angiotensin I converting enzyme showed the highest correlation with altitude in Bolivia-Brazil transect (rho = 0.45, p = 0.003). The results generate a novel hypothesis how bacterial genes can be directly involved in blood pressure regulation via rein-angiotensin system.

5.4. Discussion

We tested whether variation in altitude have effects on variation in gut microbiota of wild house mice. First, we found altitudinal differences were associated with differences in alpha- and beta-diversity measurements of the gut microbiota. Beta-diversity of the gut microbiota measured by Bray-Curtis dissimilarity was explained by differences altitude controlling for other covariates measured in this study including diet, body size, and populations. Alpha-diversity of the gut microbiota showed a weak positive correlation with altitude consistent with a previous study in humans (Lan *et al.* 2017). Especially, alpha-diversity of Bacteroidetes showed a significant correlation with altitude independent of covariates. Although altitude alone had significant effect on alpha- and beta-diversity of the gut microbiota statistically, the observed pattern can still be explained by other covariates such as climate. However, diet and host population structure are unlikely explanations of the pattern based on stable isotope measurements showing opposite correlations with altitude among the two transects (Table. S3, Fig.S1) and their colonization history (Storz *et al.* 2007). This finding is significant because correlation between altitude and the gut microbiota in humans is mostly explained by population differences including differences in diet, culture, and genetic background (Li and Zhao 2015; Li *et al.* 2016d; Lan *et al.* 2017).

To get around the issue of covariates, we took advantage of a study investigating the role of gut microbiota in sleep apnea where lower atmospheric oxygen exposure (i.e. intermittent hypoxia) lead to an enrichment of obligate anaerobes in the gut of laboratory mice compared to normoxic controls in a common environment (Moreno-Indias *et al.* 2015). This observation allowed us to test a specific hypothesis regarding whether obligate anaerobes positively correlate with altitude and other oxygen requirement types (i.e. facultative anaerobes, aerotolerant anaerobes, microaerophiles, and obligate aerobes) negatively correlate with altitude in response to atmospheric oxygen. Overall, we found 15 out of the 17 bacterial genera showing correlations with altitude in expected directions predicted by the hypothesis (Fig. 2). For example, strictly anaerobic bacteria showed positive correlations with altitude and microaerobes and aerotolerant bacteria showed negative correlations with altitude including *Akkermansia* that may even benefit from low levels of oxygen in the mucus layer (Reunanen *et al.* 2015; Ouwerkerk *et al.* 2016). The results suggest that obligate anaerobes may have greater competitive advantage at hypoxic condition compared to other oxygen requirement types as proposed by Moreno-Indias et al. (2015). The reduced partial pressure of oxygen at high altitudes lowers oxygen levels in the blood and may cause insufficient supply of oxygen to epithelial cells (Zheng, Kelly and Colgan 2015), which can result in the increased niche of the anaerobes in the luminal contents and the reduced niche of the aerobes in the mucosal surfaces as observed in mice, macaques, and humans (Espey 2013; Albenberg et al. 2014; Yasuda et al. 2015). A similar mechanism of limiting the bioavailability of the oxygen in the gut preventing the expansions of pathogenic aerobic bacteria has been described (Rivera-Chávez *et al.* 2016; Byndloss *et al.* 2017). Together, the observations suggest that the atmospheric oxygen level may affect the anaerobes and aerobes differently by directly influencing their oxygen niche (Fig. 5).

Lastly, we found that *Prevotella* was the top bacterial genera that positively correlated with altitude independently in two altitudinal transects of wild mice (Fig.3).

Prevotella is a saccharolytic anaerobe that major metabolic end products include acetic acids (Shah and Collins 1990). Interestingly, pikas (Li and Zhao 2015), ruminants (Zhang *et al.* 2016) and humans (Li *et al.* 2016d; Lan *et al.* 2017) living in high altitudes also show an enrichment of *Prevotella*. However, opposite trends between *Prevotella* and altitude has also been reported in a study of humans (Li and Zhao 2015) and macaques (Zhao *et al.* 2018). Further investigation of this relationship is necessary to understand the link. The strongest evidence of the increase in *Prevotella* due to low atmospheric oxygen comes from the hypoxic experiment in laboratory mice mentioned above where *Prevotella* was one of the few genera that showed increase in intermittent hypoxia exposure compared to normoxia controls (Moreno-Indias *et al.* 2015). Together, the repeated associations between *Prevotella* and oxygen availability across multiple mammalian species with different ecology support a possibility that atmospheric oxygen levels affecting certain members of the gut microbiota.

The high-altitude environment imposes challenges on hosts to survive and reproduce under thermoregulatory and hypoxic stress (Grindlay and Regensteiner 1983). Two major beneficial functions of the gut microbiota at high-altitude have been proposed. One is the role of the gut microbiota in increased energy harvest at high-altitude by fermenting complex carbohydrates (Li and Zhao 2015; Li *et al.* 2016d; Zhang *et al.* 2016; Lan *et al.* 2017; Zhao *et al.* 2018). A causal role of gut microbiota in increased energy harvest has been demonstrated in lab mice (Bäckhed *et al.* 2004; Turnbaugh *et al.* 2006). Many anaerobic gut bacteria produce SCFAs as end products of polysaccharide fermentation and serve as an energy source of the host (den Besten *et al.* 2013). Previous studies involving high-altitude mammals have supported this hypothesis by reporting direct and indirect evidence of greater production of SCFAs by anaerobic bacteria at higher altitudes (Li and Zhao 2015; Li *et al.* 2016d; Zhang *et al.* 2016; Lan *et al.* 2017; Zhao *et al.* 2018). Our results are also consistent with this hypothesis where we observed increase in obligate anaerobes with altitude such as *Prevotella* and unclassified genus of Lachnospiraceae. Both of them include species that produce SCFAs (Strobel 1992; Biddle *et al.* 2013) and especially *Prevotella*-dominanted gut community are known to have increased capacity to ferment polysaccharides (Kovatcheva-Datchary *et al.* 2015) and increased production of SCFAs (Chen *et al.* 2017). Predicted metagenomic functions also provide support for this hypothesis where starch and sucrose metabolism positively correlated with altitude (Fig. 4). Greater reliance on anaerobic energy harvest by the gut microbiota at high-altitudes may be beneficial to the host to conserve oxygen.

Another hypothesis is the beneficial role of the gut microbiota in the regulation of blood pressure (Pluznick 2014; Yang and Zubcevic 2017). Pluznick et al. (2014) proposed microbiota-mediated decrease and increase in blood pressure by two SCFA-receptors acting in opposite directions. For example, propionate (a type of SCFA) can act as a signaling molecule and cause a dose-dependent drop in blood pressure in mice (Pluznick *et al.* 2013) and similar vasorelaxation effects has been observed in rats (Nutting, Islam and Daugirdas 1991) and humans (Mortensen *et al.* 1990). In contrast, olfactory receptor 78 (*Olfr78*) in the kidney can increase the blood pressure by SCFA-mediated renin release (Pluznick *et al.* 2009, 2013), an enzyme that converts angiotensinogen to angiotensin I in the renin-angiotensin system. Renin-angiotensin I is further converted by angiotensin I converting enzyme to have a downstream effect of increased blood pressure by vasoconstriction (Sparks *et al.* 2014). Our results support the

involvement of the gut microbiota in blood pressure homeostasis in at least two mechanisms in the renin-angiotensin system (Fig. 5). The first mechanisms is through the activation of SCFA-receptors (Pluznick 2014) supported by the observation that high-altitude populations has greater relative abundance of SCFA-producing anaerobic bacteria compared to low altitude populations as proposed in humans (Li and Zhao 2015; Lan *et al.* 2017). The second potential mechanism is through bacterial enzymes directly affecting the renin-angiotensin system supported by the positive correlation between altitude and predicted bacterial homologs in the renin-angiotensin pathway. Especially the bacterial angiotensin I converting enzyme homolog have been shown to convert mammalian angiotensin I to angiotensin II *in vitro* (Rivière *et al.* 2007).

It is still unclear whether the observed gut microbial shifts in altitude involve genetic adaptation of the host. For example, when high-altitude adapted populations were compared with low-altitude populations at the same elevation in humans and ruminants, there were significant differences in the gut microbiota between the two groups (Qiu *et al.* 2012; Wei *et al.* 2016). This suggests that the gut microbial composition is not simply responding to the atmospheric oxygen and host genotype or dietary factors can explain this observation. In mammals, initial exposure to high altitude tends to show an increase in blood pressure (e.g. acclimation) and native populations living at high altitude tend to show reduced blood pressure compared to low-altitude populations (e.g. genetic adaptation) (Hanna 1999; Storz *et al.* 2007). Whether high-altitude adapted hosts have different mechanism controlling the gut microbiota compared to low-altitude relatives remains an open question. Further investigation of the physiological and genomic characteristics between high and low altitude populations of house mice will help distinguish the plastic and adaptive response of the gut microbiota.

In summary, we found altitude having a significant effect on the gut microbial composition of natural populations of house mice. Convergent patterns observed in the gut microbial composition across multiple species of mammals and laboratory mouse models of sleep apnea suggest that reduced partial pressure of oxygen at high altitude environment provides an advantage for obligate anaerobes compared to other oxygen requirement types. The enrichment of obligate anaerobes, SCFA-producing bacteria, and predicted metagenomic functions related to renin-angiotensin system at high altitude environment support potential roles of the gut microbiota in greater energy harvest and regulation of blood pressure in mice living in high altitudes.

5.5. Chapter 5 Figures

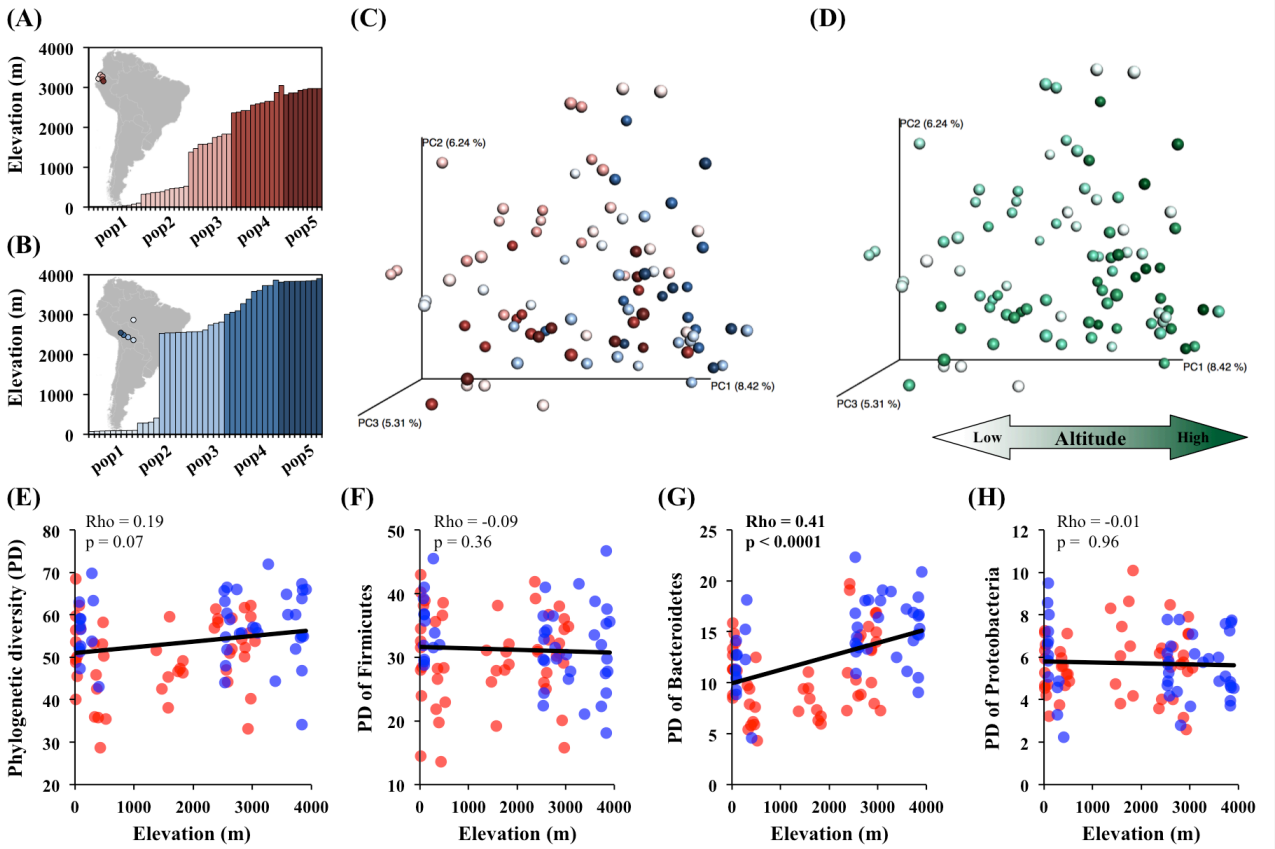


Figure 1. Effects of altitude on the gut microbiota of wild mice. Wild house mice were collected across two altitudinal gradients from Ecuador transect (n=49) **(A)** and Bolivia-Brazil transect (n=43) **(B)**. Each bar represents an individual collected from a given elevation and color coded by populations. PCoA plot of Bray-Curtis dissimilarity colored by populations (ADONIS $r^2 = 0.18$, $p<0.0001$) **(C)** and altitude (ADONIS $r^2 = 0.04$, $p<0.0001$) **(D)**. The colors of Fig.1C correspond to Fig.1A and 1B. Darker colors in Fig.1D correspond to higher altitude. A correlation between altitude and alpha-diversity measured by phylogenetic diversity (PD) **(E)**. Correlation between altitude and PD of three dominant phyla; PD of Firmicutes **(F)**, PD of Bacteroidetes **(G)**, and PD of Proteobacteria **(H)**. Only PD of Bacteroidetes significantly correlated with altitude among the dominant phyla (rho = 0.41, $p < 0.0001$). Red and blue colors correspond to individuals from Ecuador transect and Bolivia-Brazil transect, respectively.

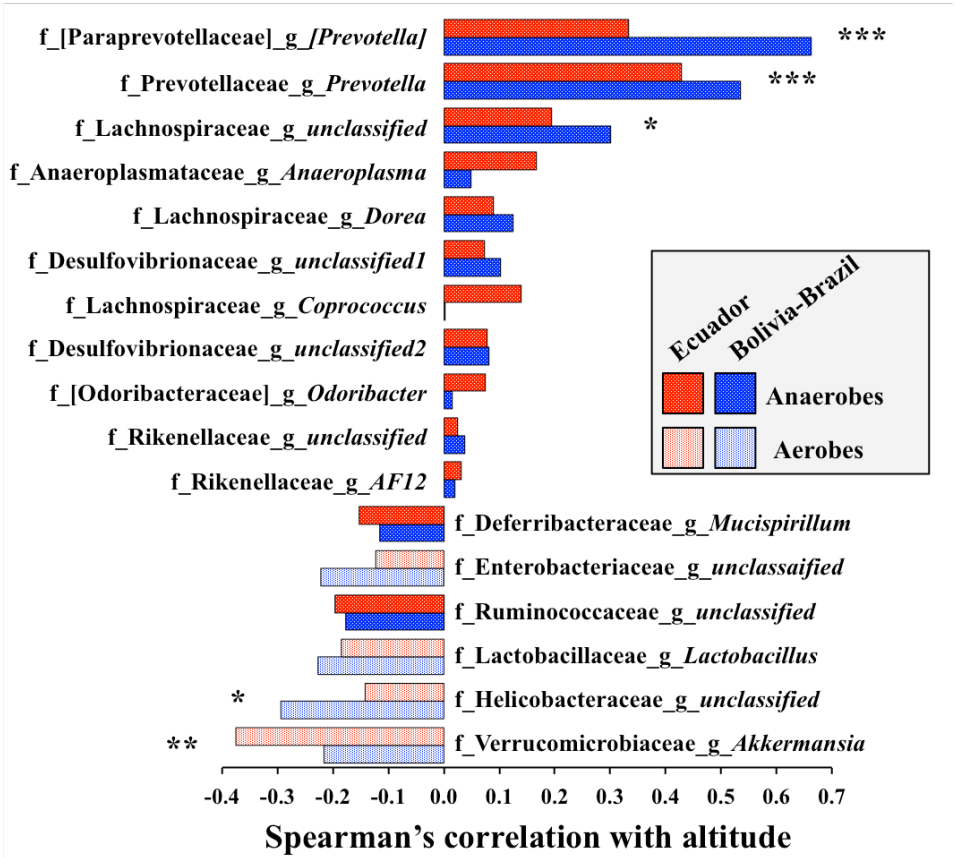


Figure 2. Correlations between altitude and bacterial genera. Bacterial genera were included in the list when (1) the correlation between altitude and relative abundance of genera was in the same direction in both transects based on Spearman's rho correlation, (2) average relative abundance >0.1% across all samples, and (3) at least named bacterial family was assigned to search for oxygen requirements. The brackets [ ] indicate recommended taxonomy. Red color indicates Ecuador transect and blue color indicate Bolivia-Brazil transect. Filled patterns show obligate anaerobes and open pattern show aerobes (i.e. facultative anaerobes, aerotolerant anaerobes, microaerophiles, and obligate aerobes). Oxygen requirements were assigned to each genera based on Bergey's Manual of Systematics of Archaea and Bacteria and recent literature. Fisher's combined p-values of the Spearman's rho raw p-values are indicated: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.0001$. After Bonferroni correction (alpha = 0.05/17 = 0.003), only the two *Prevotella* genera were significant.

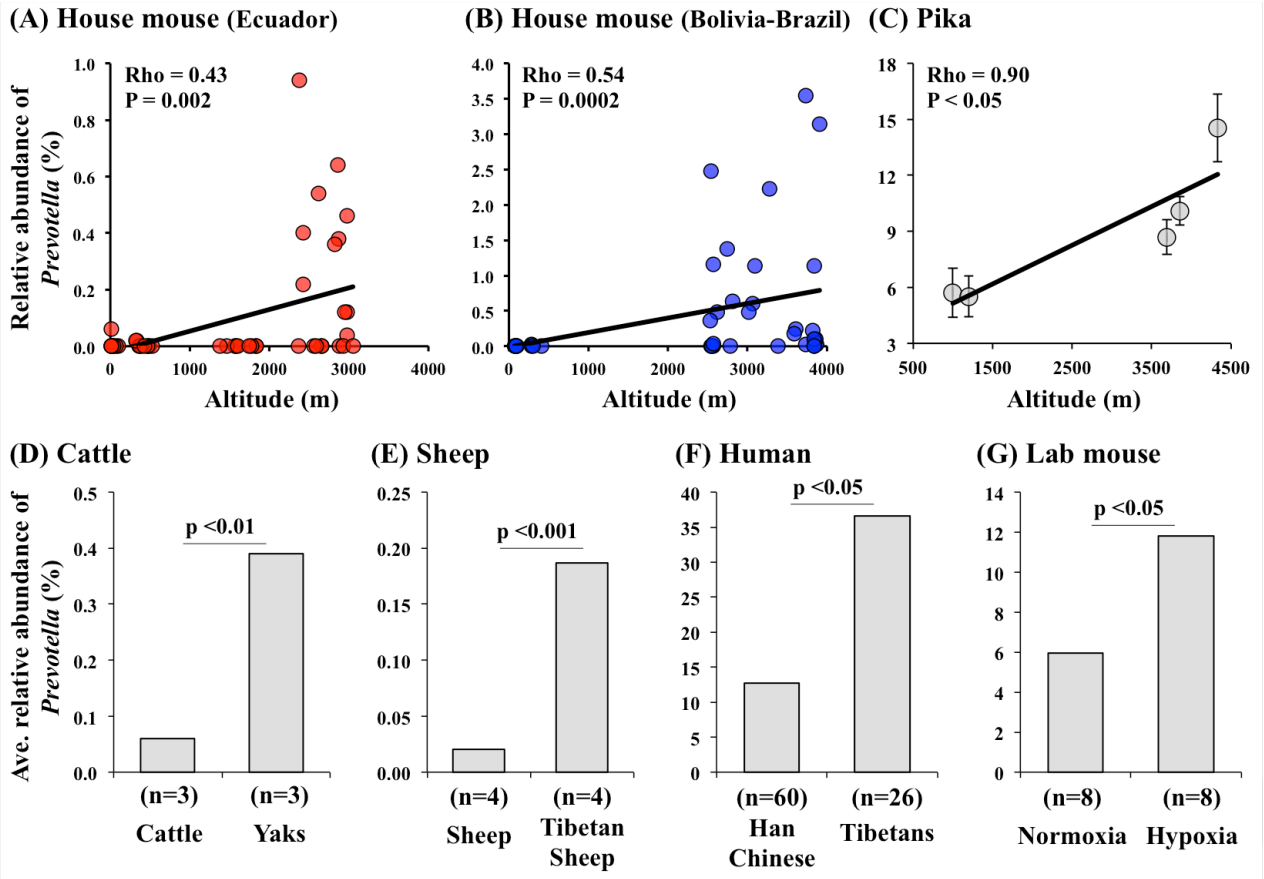


Figure 3. Convergent associations between relative abundance of *Prevotella* and high-altitude environments in different species of mammals. Significant positive correlations were observed between altitude and relative abundance of *Prevotella* in wild house mice from Ecuador transect **(A)** and Bolivia-Brazil transect **(B)**. A similar correlation between altitude and relative abundance of *Prevotella* was found in Pika (n=102) (Figure generated from data in Li et al. 2016a) **(C)**. The relative abundance of *Prevotella* was higher in Yaks compared to cattle collected from a same farm (elev. 3000m) **(D)** and in Tibetan sheep (elev. 3000m) compared to sheep (elev. 2200m) **(E)** (Figure generated from data in Zhang et al. 2016). The relative abundance of *Prevotella* was higher in Tibetans (3600-4500m) living in high altitudes compared to Han (500-3600m) living in low altitudes (Figure generated from data in Li et al. 2016b) **(F)**. In controlled lab settings, intermittent hypoxic exposure in laboratory mice resulted in higher relative abundance of *Prevotella* compared to controls (Figure generated from data in Moreno-Indias et al. 2015) **(G)**.

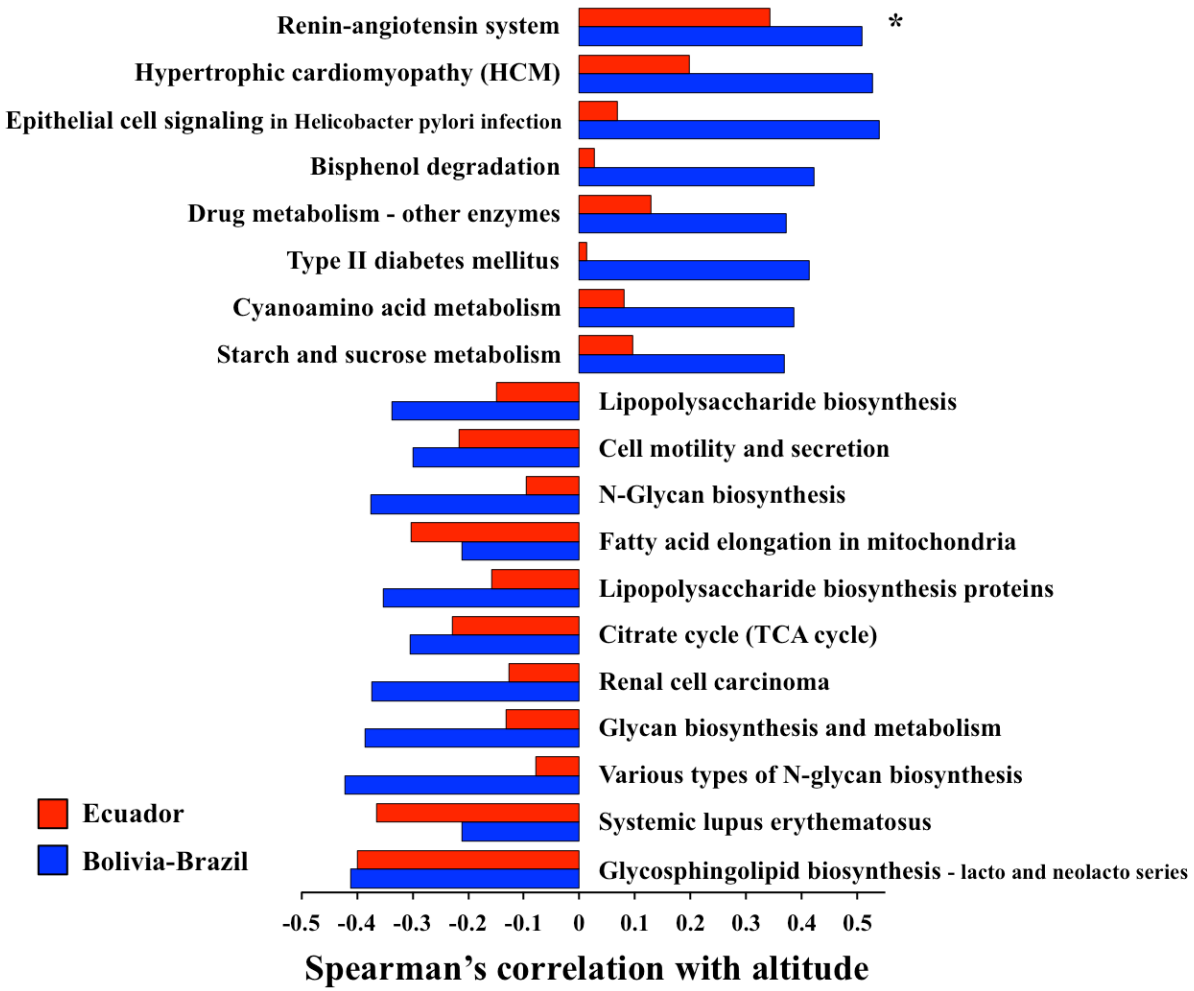


Figure 4. Correlations between altitude and predicted metagenomic functions. Spearman's correlation between altitude and 19 KEGG pathways (level 3) that show Fisher's combined p-value < 0.05 are shown. After correcting for multiple testing (Bonferroni correction; alpha = 0.05/183 = 0.0003) only Renin-angiotensin system remained significant (Ecuador; rho = 0.34, p = 0.0159. Bolivia-Brazil; rho = 0.51, p-value = 0.005. Fisher's combined p-value = 0.0001). Spearman's rho correlation for Ecuador transect is shown in red, and Bolivia-Brazil transect is shown in blue.

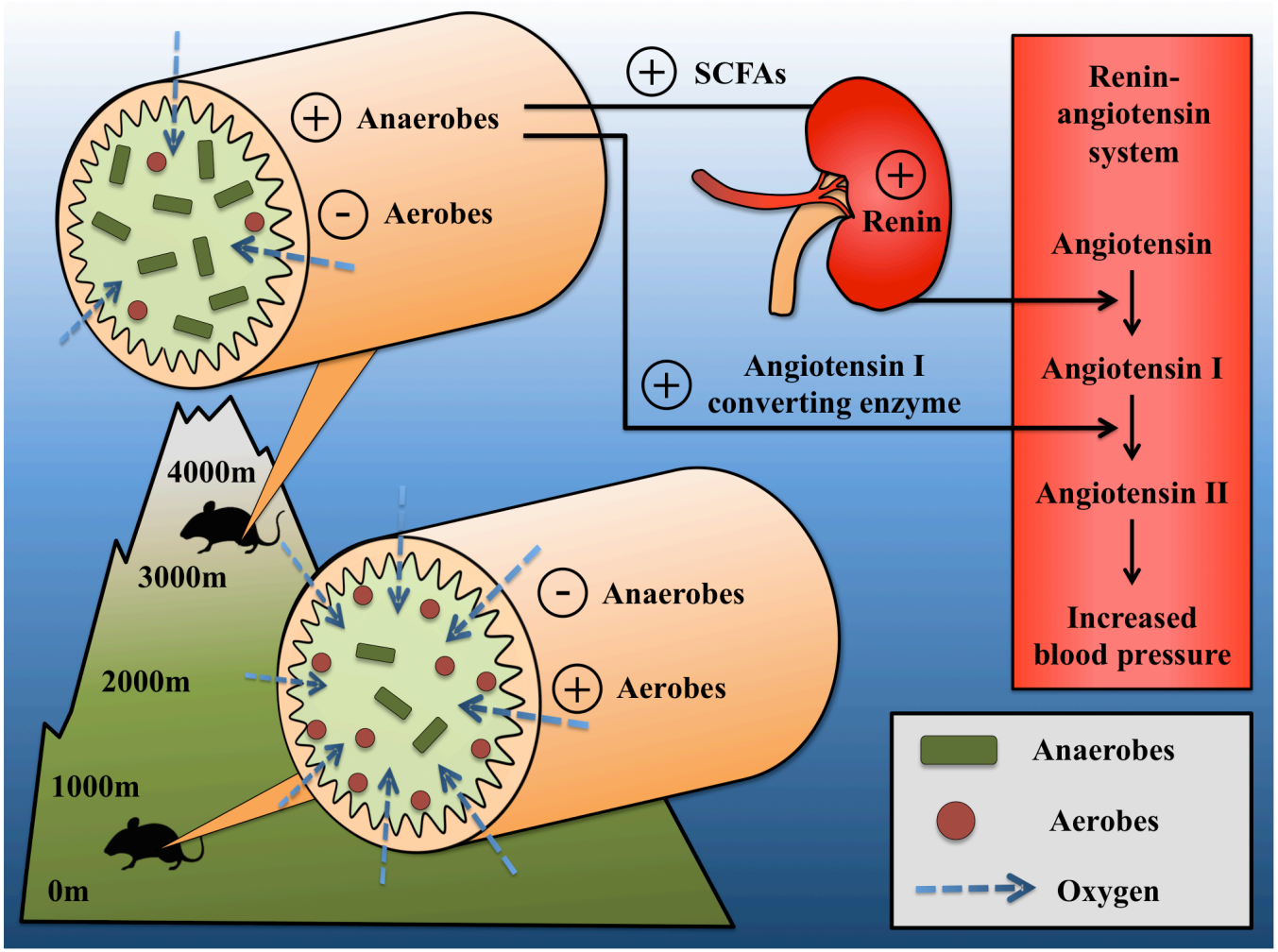


Figure 5. Proposed mechanism of the microbiota-mediated regulation of blood pressure in response to atmospheric oxygen at high altitude. Anaerobes refer to obligate anaerobes and aerobes refer to all other oxygen requirements (i.e. facultative anaerobes, aerotolerant anaerobes, microaerophiles, and obligate aerobes).

5.6. Supplemental Information

Table S1. Sample information (continued).

ID	Alt. (m)	Lat.	Long.	Country	Locality	Sex	Reproductive state	Sample storage	Weight	BMI	Carbon	Nitrogen	Climate PC1	Climate PC2
FMM54	10	-0.8461	-80.1649	Ecuador	Manabi	M	na	RNAlater	13.0	0.00219	-21.7	7.2	2.617	2.085
FMM66	14	-0.9254	-80.4793	Ecuador	Manabi	F	non-pregnant	RNAlater	9.5	0.00152	-13.2	8.5	1.493	2.791
FMM56	15	-0.8549	-80.1629	Ecuador	Manabi	F	non-pregnant	RNAlater	13.0	0.00144	-10.7	9	2.617	2.085
FMM63	16	-0.8483	-80.1634	Ecuador	Manabi	M	na	RNAlater	7.5	0.00123	-21.8	4.6	2.617	2.085
FMM61	16	-0.8503	-80.1640	Ecuador	Manabi	F	pregnant	RNAlater	13.5	0.00240	-24.7	6	2.617	2.085
FMM64	16	-0.8503	-80.1640	Ecuador	Manabi	M	na	RNAlater	9.3	0.00178	-8.6	7.6	2.617	2.085
FMM51	18	-0.8830	-80.1341	Ecuador	Manabi	M	na	RNAlater	14.0	0.00249	-11.8	8.7	2.617	2.085
FMM55	33	-0.8567	-80.1646	Ecuador	Manabi	F	non-pregnant	RNAlater	11.0	0.00238	-13.7	21.1	2.617	2.085
FMM60	43	-1.0554	-80.4901	Ecuador	Manabi	M	na	RNAlater	9.8	0.00164	-18.1	8.7	1.521	2.673
FMM58	73	-1.0808	-80.5233	Ecuador	Manabi	F	non-pregnant	RNAlater	17.3	0.00208	-10.3	11.9	1.140	2.648
FMM48	100	-1.0794	-80.5399	Ecuador	Manabi	M	na	RNAlater	11.0	0.00176	-16.2	8.4	1.140	2.648
FMM94	320	-0.2439	-79.3369	Ecuador	Santo_Domingo	F	non-pregnant	RNAlater	4.0	0.00119	-15	7.9	5.310	-0.048
FMM104	335	-0.2442	-79.3385	Ecuador	Santo_Domingo	M	na	RNAlater	9.5	0.00148	-18	7.7	5.310	-0.048
FMM100	361	-0.2443	-79.3274	Ecuador	Santo_Domingo	F	non-pregnant	RNAlater	15.0	0.00185	-18.4	5.6	5.316	-0.451
FMM92	369	-0.1249	-79.2584	Ecuador	Santo_Domingo	F	pregnant	RNAlater	25.0	0.00372	-22.1	8.7	6.104	-0.946
FMM91	388	-0.1173	-79.2591	Ecuador	Santo_Domingo	M	na	RNAlater	9.8	0.00164	-15.8	9	6.104	-0.946
FMM98	429	-0.2440	-79.2613	Ecuador	Santo_Domingo	F	non-pregnant	RNAlater	10.3	0.00093	-22.4	8	5.316	-0.451
FMM88	465	-0.2068	-79.1809	Ecuador	Santo_Domingo	F	pregnant	RNAlater	9.0	0.00131	-21.1	8.2	5.316	-0.451
FMM103	477	-0.1897	-79.3184	Ecuador	Santo_Domingo	M	na	RNAlater	13.5	0.00196	-20.1	7.5	5.316	-0.451
FMM105	487	-0.2796	-79.2118	Ecuador	Santo_Domingo	M	na	RNAlater	12.8	0.00161	-14.7	7.6	5.316	-0.451
FMM110	525	-0.3910	-79.2102	Ecuador	Santo_Domingo	M	na	RNAlater	7.0	0.00112	-11.9	6.6	4.944	-0.355
FMM70	1377	no data	no data	Ecuador	Pichincha_Nanegalito	F	pregnant	RNAlater	20.0	0.00247	-16.8	6.4	no data	no data
FMM69	1470	-0.0330	-78.6814	Ecuador	Pichincha_Nanegalito	M	na	RNAlater	11.5	0.00171	-17.5	7	2.481	-2.858
FMM67	1577	-0.0118	-78.6725	Ecuador	Pichincha_Nanegalito	M	na	RNAlater	13.5	0.00153	-20.9	8.6	2.481	-2.858
FMM68	1577	-0.0118	-78.6681	Ecuador	Pichincha_Nanegalito	M	na	RNAlater	13.3	0.00147	-21.8	8.7	2.481	-2.858
FMM75	1599	-0.0622	-78.6820	Ecuador	Pichincha_Nanegalito	F	pregnant	RNAlater	22.0	0.00266	-18.1	8.3	2.481	-2.858
FMM84	1747	-0.0760	-78.6960	Ecuador	Pichincha_Nanegalito	M	na	RNAlater	12.3	0.00113	-14.9	8.6	2.481	-2.858
FMM83	1774	-0.0649	-78.6893	Ecuador	Pichincha_Nanegalito	M	na	RNAlater	10.3	0.00127	-19.2	7	2.481	-2.858
FMM79	1832	-0.0369	-78.6968	Ecuador	Pichincha_Nanegalito	F	pregnant	RNAlater	-	-	-16.6	7.4	2.481	-2.858
FMM81	1832	-0.0369	-78.6968	Ecuador	Pichincha_Nanegalito	M	na	RNAlater	4.0	0.00087	-17.1	7.3	2.481	-2.858
FMM6	2363	-0.1975	-78.3941	Ecuador	Pichincha_Tumbaco	M	na	RNAlater	15.0	0.00213	-13	6.4	-0.662	-1.553
FMM34	2381	-0.2116	-78.3870	Ecuador	Pichincha_Tumbaco	F	pregnant	RNAlater	13.0	0.00198	-15	6.8	-0.662	-1.553

Table S1. Sample information (continued).

ID	Alt. (m)	Lat.	Long.	Country	Locality	Sex	Reproductive state	Sample storage	Weig ht	BMI	Carbon	Nitrogen	Climate PC1	Climate PC2
FMM16	2421	-0.8736	-78.6071	Ecuador	Pichincha_Tumbaco	F	pregnant	RNAlater	10.5	0.00177	-9.0	7.7	-1.970	-2.525
FMM17	2421	-0.8736	-78.6071	Ecuador	Pichincha_Tumbaco	F	pregnant	RNAlater	-	-	-8.8	7.8	-1.970	-2.525
FMM35	2557	-0.1764	-78.3289	Ecuador	Pichincha_Tumbaco	M	na	RNAlater	17.0	0.00192	-16.2	9.0	-1.021	-5.760
FMM9	2589	-0.2479	-78.3570	Ecuador	Pichincha_Tumbaco	F	pregnant	RNAlater	12.0	0.00192	-21.0	7.3	-0.662	-1.553
FMM12	2614	-0.2482	-78.3552	Ecuador	Pichincha_Tumbaco	M	na	RNAlater	15.0	0.00306	-11.0	7.6	-0.662	-1.553
FMM13	2653	-0.2415	-78.3350	Ecuador	Pichincha_Tumbaco	M	na	RNAlater	13.8	0.00204	-9.4	8.7	-0.662	-1.553
FMM14	2653	-0.2528	-78.3387	Ecuador	Pichincha_Tumbaco	F	non-pregnant	RNAlater	13.3	0.00188	-10.8	8.3	-0.662	-1.553
FMM33	2818	-0.8561	-78.6132	Ecuador	Cotopaxi	F	non-pregnant	RNAlater	13.5	0.00216	-16.7	6.3	-1.970	-2.525
FMM25	2865	-0.8697	-78.6214	Ecuador	Cotopaxi	F	non-pregnant	RNAlater	11.0	0.00168	-12.4	8.6	-1.970	-2.525
FMM24	2867	-0.8633	-78.6119	Ecuador	Cotopaxi	F	pregnant	RNAlater	13.5	0.00183	-9.5	9.6	-1.970	-2.525
FMM15	2875	-0.8550	-78.6091	Ecuador	Pichincha_Tumbaco	M	na	RNAlater	13.5	0.00187	-10.5	7.6	-1.970	-2.525
FMM30	2927	-0.8366	-78.6632	Ecuador	Cotopaxi	M	na	RNAlater	6.5	0.00141	-20.5	8.8	-1.970	-2.525
FMM31	2951	-0.8268	-78.6670	Ecuador	Cotopaxi	M	na	RNAlater	21.5	0.00260	-20.2	7.8	-2.317	-3.380
FMM18	2973	-0.8736	-78.6071	Ecuador	Cotopaxi	M	na	RNAlater	12.5	0.00177	-21.5	11.4	-1.970	-2.525
FMM19	2973	-0.8736	-78.6071	Ecuador	Cotopaxi	F	non-pregnant	RNAlater	11.5	0.00171	-17.9	9.2	-1.970	-2.525
FMM32	2973	-0.8736	-78.6071	Ecuador	Cotopaxi	M	na	RNAlater	14.0	0.00219	-17.0	8.2	-1.970	-2.525
FMM38	3050	-0.4292	-78.4315	Ecuador	Pichincha_Tumbaco	F	pregnant	RNAlater	19.0	0.00276	-17.1	9.3	-0.938	-3.627
FMM.270	72	-8.7779	-63.8415	Brazil	Porto_Velho	F	pregnant	Flash frozen	15.0	0.00163	-16.3	8.1	3.796	1.067
FMM.271	77	-8.7645	-63.8398	Brazil	Porto_Velho	F	non-pregnant	Flash frozen	12.3	0.00162	-24.0	9.9	3.796	1.067
FMM.269	83	-8.7825	-63.8457	Brazil	Porto_Velho	F	pregnant	Flash frozen	13.8	0.00166	-20.0	10.1	3.796	1.067
FMM.279	86	-8.7763	-63.8004	Brazil	Porto_Velho	F	pregnant	Flash frozen	14.0	0.00158	-23.6	10.6	3.501	1.339
FMM.277	88	-8.7828	-63.8520	Brazil	Porto_Velho	M	na	Flash frozen	7.0	0.00115	-16.9	8.4	3.796	1.067
FMM.278	88	-8.7843	-63.8033	Brazil	Porto_Velho	F	non-pregnant	Flash frozen	13.0	0.00180	-16.2	7.7	3.501	1.339
FMM.266	92	-8.7715	-63.8029	Brazil	Porto_Velho	F	non-pregnant	Flash frozen	12.5	0.00158	-18.4	9.4	3.501	1.339
FMM.267	94	-8.7841	-63.8416	Brazil	Porto_Velho	M	na	Flash frozen	10.0	0.00145	-20.1	9.8	3.796	1.067
FMM.268	97	-8.7716	-63.7976	Brazil	Porto_Velho	M	na	Flash frozen	12.3	0.00178	-20.0	9.5	3.501	1.339
TAS595	281	-17.3280	-63.2565	Bolivia	Santa_Cruz	M	na	RNAlater	12.9	0.00187	-26.0	12.8	1.940	1.686
TAS594	282	-17.3268	-63.2571	Bolivia	Santa_Cruz	F	non-pregnant	RNAlater	13.8	0.00191	-	-	1.940	1.686
TAS598	306	-17.3306	-63.2491	Bolivia	Santa_Cruz	F	pregnant	RNAlater	14.5	0.00181	-25.9	13.7	1.940	1.686
TAS592	409	-17.7348	-63.1696	Bolivia	Santa_Cruz	F	non-pregnant	RNAlater	14.2	0.00192	-25.3	13.3	2.113	1.184
TAS576	2530	-17.4705	-66.3413	Bolivia	Cochabamba	F	pregnant	RNAlater	25.5	0.00245	-10.6	12.8	-3.227	1.327
TAS578	2542	-17.4371	-66.3341	Bolivia	Cochabamba	M	na	RNAlater	18.6	0.00215	-16.7	9.2	-3.227	1.327
TAS579	2542	-17.4276	-66.3312	Bolivia	Cochabamba	F	non-pregnant	RNAlater	18.2	0.00215	-20.9	13.8	-2.653	3.072

Table S1. Sample information (continued).

ID	Alt. (m)	Lat.	Long.	Country	Locality	Sex	Reproductive state	Sample storage	Weight	BMI	Carbon	Nitrogen	Climate PC1	Climate PC2
TAS588	2551	-17.4365	-66.1618	Bolivia	Cochabamba	M	na	RNAlater	17.0	0.00197	-18.6	10.4	-2.834	2.159
TAS580	2552	-17.3956	-66.3123	Bolivia	Cochabamba	M	na	RNAlater	18.2	0.00210	-16.9	11.1	-2.653	3.072
TAS574	2569	-17.3911	-66.2303	Bolivia	Cochabamba	F	pregnant	RNAlater	15.2	0.00210	-12.7	9.6	-2.653	3.072
TAS589	2571	-17.4464	-66.1680	Bolivia	Cochabamba	F	non-pregnant	RNAlater	17.8	0.00168	-10.9	16.1	-2.653	3.072
TAS582	2575	-17.3795	-66.3066	Bolivia	Cochabamba	F	non-pregnant	RNAlater	15.6	0.00197	-15.2	12.3	-2.653	3.072
TAS573	2615	-17.3738	-66.1528	Bolivia	Cochabamba	M	na	RNAlater	16.7	0.00185	-15.6	7.5	-2.834	2.159
TAS587	2740	-17.5397	-66.0132	Bolivia	Cochabamba	M	na	RNAlater	17.1	0.00207	-10.9	8.8	-2.903	2.302
TAS586	2783	-17.6170	-66.0169	Bolivia	Cochabamba	M	na	RNAlater	15.6	0.00201	-19.2	10.7	-2.903	2.302
TAS585	2815	-17.6442	-65.9884	Bolivia	Cochabamba	M	na	RNAlater	11.8	0.00176	-21.1	9.1	-2.865	2.628
TAS572	3011	-16.6281	-68.0517	Bolivia	La_Paz	F	non-pregnant	RNAlater	11.3	0.00177	-21.7	8.7	-4.102	0.365
TAS554	3060	-16.6022	-68.0651	Bolivia	La_Paz	F	non-pregnant	RNAlater	9.3	0.00179	-21.9	16.0	-4.102	0.365
TAS565	3097	-16.5892	-68.0691	Bolivia	La_Paz	M	na	RNAlater	24.3	0.00243	-20.2	10.7	-4.102	0.365
TAS555	3276	-16.5742	-68.0786	Bolivia	La_Paz	M	na	RNAlater	17.8	0.00220	-20.3	8.2	-4.102	0.365
TAS548	3388	-16.5391	-68.0710	Bolivia	La_Paz	F	non-pregnant	RNAlater	17.0	0.00227	-20.5	8.9	-4.102	0.365
TAS571	3583	-16.5806	-68.1274	Bolivia	La_Paz	M	na	RNAlater	13.6	0.00164	-15.5	15.3	-4.102	0.365
TAS570	3607	-16.5841	-68.1308	Bolivia	La_Paz	F	non-pregnant	RNAlater	16.3	0.00220	-17.4	7.6	-4.102	0.365
TAS567	3728	-16.5842	-68.1513	Bolivia	La_Paz	M	na	RNAlater	19.0	0.00213	-22.4	10.4	-4.102	0.365
TAS568	3731	-16.5880	-68.1521	Bolivia	La_Paz	F	non-pregnant	RNAlater	12.3	0.00202	-20.5	7.8	-4.102	0.365
TAS611	3816	-16.2531	-68.5672	Bolivia	Lake_Titikaka	M	na	RNAlater	16.6	0.00188	-22.4	7.4	-4.545	-0.108
TAS610	3835	-16.2021	-68.5888	Bolivia	Lake_Titikaka	F	non-pregnant	RNAlater	10.8	0.00165	-22.8	8.6	-4.545	-0.108
TAS613	3835	-16.1836	-68.7684	Bolivia	Lake_Titikaka	M	na	RNAlater	17.5	0.00219	-25.4	5.4	-4.325	-0.051
TAS601	3838	-16.1902	-68.6007	Bolivia	Lake_Titikaka	M	na	RNAlater	20.1	0.00238	-23.6	6.2	-4.545	-0.108
TAS612	3839	-16.2757	-68.5511	Bolivia	Lake_Titikaka	M	na	RNAlater	10.9	0.00179	-21.8	9.6	-4.545	-0.108
TAS600	3845	-16.3326	-68.8250	Bolivia	Lake_Titikaka	M	na	RNAlater	13.8	0.00191	-22.4	8.3	-4.325	-0.051
TAS599	3855	-16.1716	-68.8284	Bolivia	Lake_Titikaka	F	pregnant	RNAlater	15.8	0.00209	-24.1	7.1	-4.325	-0.051
TAS549	3866	-16.5477	-68.0251	Bolivia	La_Paz	M	na	RNAlater	14.6	0.00193	-21.1	8.8	-4.102	0.365
TAS609	3906	-16.2200	-68.5830	Bolivia	Lake_Titikaka	F	non-pregnant	RNAlater	17.6	0.00215	-22.0	9.0	-4.545	-0.108

Table S2. Correlations between Bray-Curtis dissimilarity and predictor variables using ADONIS.

Variables[a]	All samples			Ecuador			Bolivia-Brazil		
	n	R^2	p-value[b]	n	R^2	p-value*	n	R^2	p-value*
Altitude	92	**0.038**	**<.0001**	49	0.032	0.021	43	**0.081**	**<.0001**
Body weight	90	0.017	0.021	47	0.025	0.192	43	0.035	0.022
BMI	90	0.015	0.057	47	0.023	0.357	43	**0.049**	**0.0002**
Diet (Carbon)	91	0.018	0.003	49	0.033	0.014	42	0.032	0.054
Diet (Nitrogen)	91	0.013	0.135	49	0.017	0.817	42	0.027	0.232
Population[1]	92	**0.183**	**<.0001**	49	0.116	0.003	43	**0.194**	**<.0001**
Pregnancy[2]	45	0.020	0.721	23	0.046	0.409	22	0.042	0.694
Sex	92	0.009	0.821	49	0.019	0.671	43	0.021	0.691

[a] Altitude, Body weight, Body mass index (BMI), and Carbon and Nitrogen stable isotope diet measurements are continuous variables. Population (five populations for each altitudinal transect), Pregnancy (pregnant vs non-pregnant), and Sex (female vs male) are categorical variables.

* Significant p-values after Bonferroni correction are bolded (alpha = 0.05/24 = 0.0021)

Table S3. Correlations between altitude and metadata.

	All samples (n=92)		Ecuador (n=49)		Bolivia-Brasil (n=43)	
	rho	p-value	rho	p-value	rho	p-value
Body weight	**0.3448**	**0.0009**	0.2716	0.0648	0.243	0.1163
BMI	**0.3327**	**0.0014**	0.1729	0.2451	**0.4768**	**0.0012**
Carbon	-0.1888	0.0731	0.0982	0.5021	-0.2875	0.0648
Nitrogen	0.0314	0.7678	0.1392	0.3402	-0.3706	0.0157

Raw p-values are shown. Significance after Bonferroni correction are bolded (alpha = 0.05/12 = 0.0042)

Table S4. Correlations between altitude and alpha-diversity measurements.

	All samples		Ecuador		Bolivia-Brazil	
	rho	p-value*	rho	p-value*	rho	p-value*
OTU counts	0.09	0.39	0.13	0.37	-0.12	0.44
Phylogenetic diversity	0.19	*0.07*	0.12	0.42	0.14	0.37
Simpson	0.23	*0.03*	0.11	0.46	0.21	0.18

*Raw p-values are shown. None of the correlations are significant after Bonferroni correction (alpha = 0.05/9 = 0.006).

Table S5. Correlations between altitude and relative abundances of bacterial phyla.

Phyla[1]	Average relative abudance	Ecuador		Bolivia-Brazil		Fisher's combined p-value [2]
		rho	p-value	rho	p-value	
Firmicutes	0.44	0.13	0.36	0.17	0.29	0.34
Bacteroidetes	0.33	-0.06	0.69	0.14	0.36	-
Proteobacteria	0.17	-0.10	0.51	-0.26	*0.09*	0.19
Deferribacteres	0.04	-0.15	0.30	-0.12	0.46	0.41
Unclassified Phylum	0.01	0.05	0.73	0.18	0.24	0.48
Tenericutes	0.004	0.01	0.95	0.11	0.49	0.82
Verrucomicrobia	0.003	-0.05	0.73	0.15	0.35	-
Fusobacteria	0.002	-0.11	0.44	0.15	0.32	-
Cyanobacteria	0.002	**0.29**	**0.04**	**0.31**	**0.04**	0.01
Actinobacteria	0.001	0.03	0.83	0.26	0.10	0.28

[1] Bacterial phyla that has average relative abundance >0.1% across all samples were included.

[2] None of the combined p-values are significant after Bonferroni correction (alpha = 0.05/7 = 0.007).

Table S6. Correlations between altitude and the relative abundances of 23 bacterial genera with their oxygen requirements (continued).

Bacterial genera that correlated with altitude[1]			Ecuador		Bolivia-Brazil		Fisher's combined p-value	Oxygen requirements[2]	References
Phyla	Family	Genera	rho	p-value	rho	p-value			
Bacteroidetes	[Paraprevotellaceae]	*[Prevotella]*	**0.33**	**0.02**	**0.66**	**<.0001**	**<.0001**	Obligate anaerobes	Bergey's mannual
Bacteroidetes	Prevotellaceae	*Prevotella*	**0.43**	**0.002**	**0.54**	**0.0002**	**<.0001**	Obligate anaerobes	Bergey's mannual
Proteobacteria	Burkholderiales_unc	*Burkholderiales_unc*	0.14	0.32	**0.55**	**0.0001**	**0.0004**	*unclassified*	-
Verrucomicrobia	Verrucomicrobiaceae	*Akkermansia*	**-0.38**	**0.008**	-0.22	0.16	**0.01**	Oxygen tolerant	Reunanen et al. 2015, Ouwerkerk et al. 2016
Firmicutes	Lachnospiraceae	*Lachnospiraceae_unc*	0.19	0.18	0.30	0.05	0.05	Obligate anaerobes	Bergey's mannual
Proteobacteria	Helicobacteraceae	*Helicobacteraceae_unc*	-0.14	0.33	-0.29	0.06	0.09	Microaerobes*	Bergey's mannual
Cyanobacteria	YS2_unc	*YS2_unc*	0.22	0.14	0.22	0.16	0.11	*uncultured*	-
Firmicutes	Lactobacillaceae	*Lactobacillus*	-0.18	0.20	-0.23	0.14	0.13	Facultative anaerobes	Bergey's mannual
Firmicutes	Ruminococcaceae	*Ruminococcaceae_unc*	-0.20	0.18	-0.18	0.26	0.19	Obligate anaerobes	Bergey's mannual
Proteobacteria	Enterobacteriaceae	*Enterobacteriaceae_unc*	-0.12	0.40	-0.22	0.15	0.23	Facultative anaerobes	Bergey's mannual
Bacteroidetes	Bacteroidales_unc	*Bacteroidales_unc*	-0.08	0.59	-0.22	0.15	0.30	*unclassified*	-
Deferribacteres	Deferribacteraceae	*Mucispirillum*	-0.15	0.30	-0.12	0.46	0.41	Obliate anaerobes	Robertson et al. 2005

[1] Bacterial genera were included in the list when (1) the correlation between altitude and relative abudance of taxa was in the same direction across the two mountains based on Spearman's rho correlation and (2) average relative abundance >0.1% across all samples. The brackets [ ] indicate recommended taxonomy.

[2] Oxygen reuqirements were assigned based on Bergey's Manual of Systematics of Archea and Bacteria and recent liteature. When the genera were unclassified, we used the oxygen requirements of the family. When the family of the unclassified genera included obligate anaerobes and all other oxygen requrement types, we searched for all the recognized genera within the family and assigned oxygen requirement based on majority rule (i.e. two out of the three genera showed the same oxygen requirements) in all such cases indicated by *.

Table S6. Correlations between altitude and the relative abundances of 23 bacterial genera with their oxygen requirements (continued).

Bacterial genera that correlated with altitude[1]			Ecuador		Bolivia-Brazil		Fisher's combined p-value	Oxygen requirements[2]	References
Phyla	Family	Genera	rho	p-value	rho	p-value			
Firmicutes	Clostridiales_unc	*Clostridiales_unc*	0.10	0.48	0.15	0.34	0.46	*unclassified*	-
Tenericutes	Anaeroplasmataceae	*Anaeroplasma*	0.17	0.25	0.05	0.75	0.51	Obligate anaerobes	Bergey's mannual
Firmicutes	Lachnospiraceae	*Dorea*	0.09	0.54	0.13	0.42	0.57	Obligate anaerobes	Bergey's mannual
Proteobacteria	Desulfovibrionaceae	*Desulfovibrionaceae_unc1*	0.07	0.62	0.10	0.51	0.68	Obligate anaerobes*	Bergey's mannual
Firmicutes	Lachnospiraceae	*Coprococcus*	0.14	0.34	0.00	1.00	0.70	Obligate anaerobes	Bergey's mannual
Proteobacteria	Desulfovibrionaceae	*Desulfovibrionaceae_unc2*	0.08	0.59	0.08	0.60	0.72	Obligate anaerobes*	Bergey's mannual
Firmicutes	Clostridiales_unc	*Clostridiales_unc*	-0.10	0.49	-0.05	0.73	0.73	*unclassified*	-
Bacteroidetes	S24-7	*S24-7_unc*	0.08	0.58	0.06	0.70	0.77	*uncultured*	-
Bacteroidetes	[Odoribacteraceae]	*Odoribacter*	0.07	0.61	0.02	0.92	0.89	Obligate anaerobes	Hardham et al. 2008
Bacteroidetes	Rikenellaceae	*Rikenellaceae_unc*	0.03	0.86	0.04	0.81	0.95	Obligate anaerobes	Bergey's mannual
Bacteroidetes	Rikenellaceae	*AF12*	0.03	0.83	0.02	0.90	0.97	Obligate anaerobes	Bergey's mannual

[1] Bacterial genera were included in the list when (1) the correlation between altitude and relative abudance of taxa was in the same direction across the two mountains based on Spearman's rho correlation and (2) average relative abundance >0.1% across all samples. The brackets [ ] indicate recommended taxonomy.

[2] Oxygen reuqirements were assigned based on Bergey's Manual of Systematics of Archea and Bacteria and recent liteature. When the genera were unclassified, we used the oxygen requirements of the family. When the family of the unclassified genera included obligate anaerobes and all other oxygen requrement types, we searched for all the recognized genera within the family and assigned oxygen requirement based on majority rule (i.e. two out of the three genera showed the same oxygen requirements) in all such cases indicated by *.

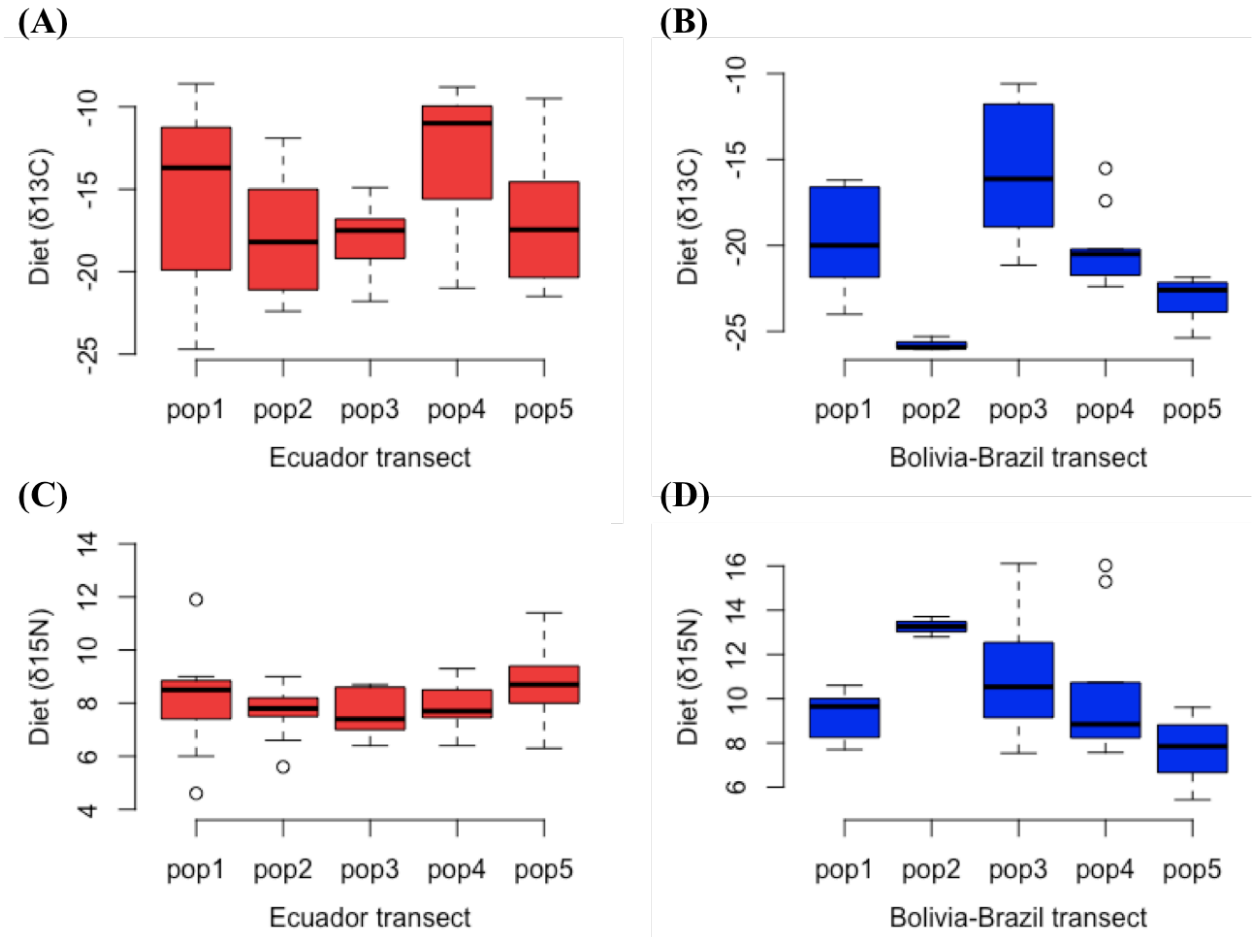


Figure S1. Box plots of carbon and nitrogen stable isotope diet measurements. Carbon isotope measurements differed by population in Ecuador transect (ANOVA $p = 0.04$) (A) and Bolivia-Brazil transect (ANOVA $p < 0.0001$). Nitrogen isotope measurements did not differ by population in Ecuador transect (ANOVA $p = 0.4$) (C), but did vary in Bolivia-Brazil transect (ANOVA $p = 0.003$) (D). The two transects significantly differ in their carbon (ANOVA $p < 0.0001$) and nitrogen (ANOVA $p = 0.0009$) stable isotope measurements.

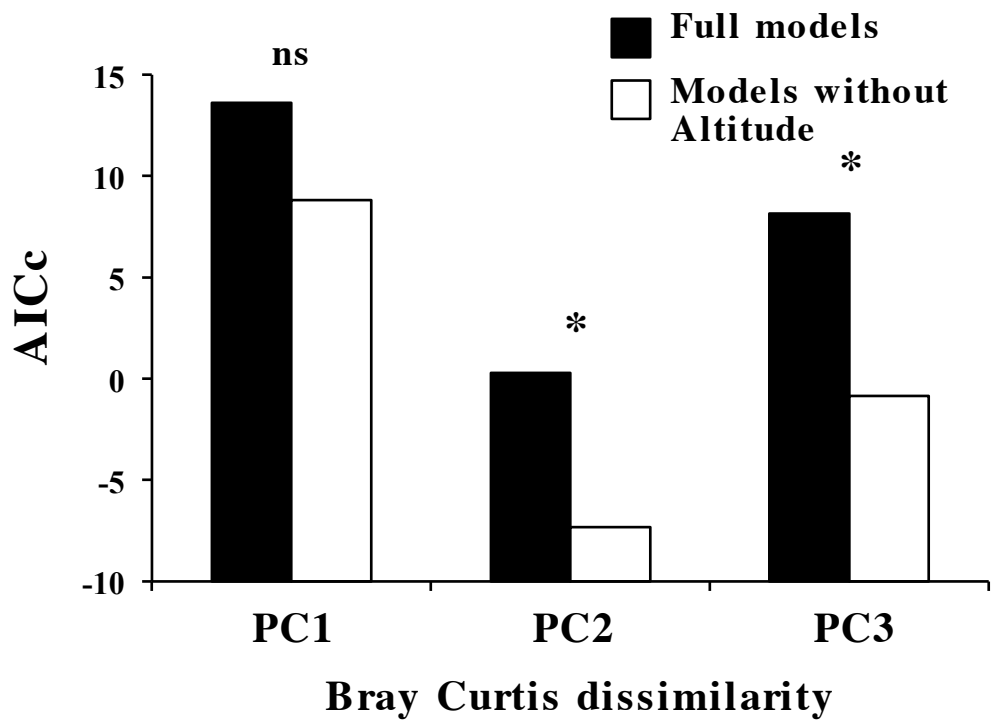


Figure S2. Model comparisons using Linear Mixed-Effects models. The response variables were Bray-Curtis dissimilarity PC1 (8.4%), PC2 (6.2%), and PC3 (5.3%). The full model included five fixed effects (altitude, body weight, BMI, carbon, and nitrogen) and three random effects (population, reproductive status, and sample storage method). The full models were compared to models without altitude using Akaike information criterion with sample size correction (AICc). Significance is based on likelihood ratio test p-values; * $p < 0.05$, ns $p > 0.05$.

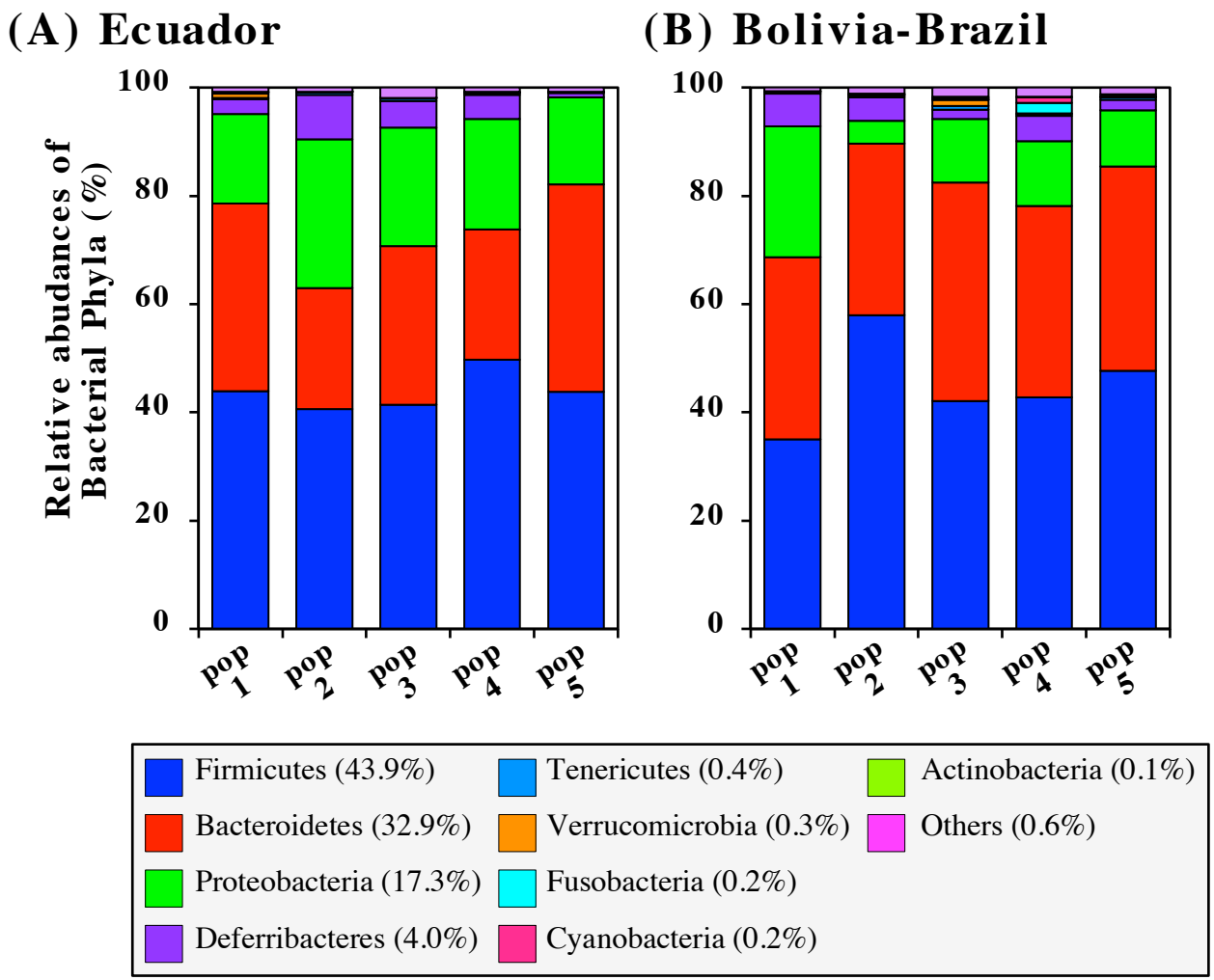


Figure S3. Average relative abundance of bacterial phyla per population. Pop1-5 are in the order of low to high altitude in Ecuador transect (A) and Bolivia-Brazil transect (B). The colors correspond to bacterial phyla that showed average relative abundance greater than 0.1% across all samples.

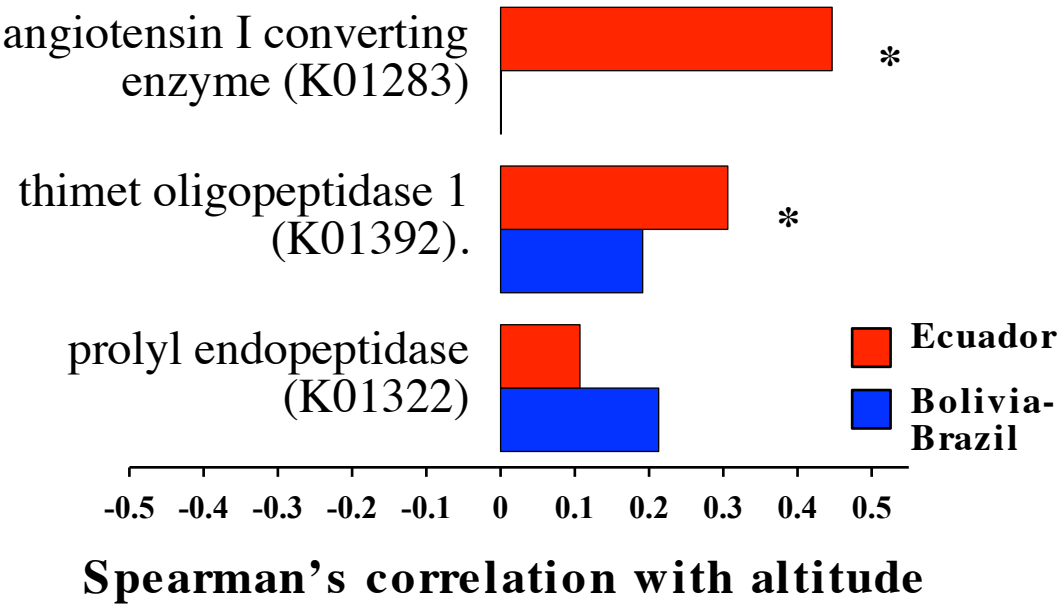


Figure S4. Correlations between altitude and predicted KEGG orthologs of renin-angiotensin system (ko04614). Spearman's correlation between altitude and three KEGG orthologs are shown for both Ecuador transect and Bolivia-Brazil transect. Spearman's correlation with Fisher's combined p-value < 0.05 are indicated by *.

Chapter 6

The gut microbiota and Bergmann's rule in wild mice

Anticipated co-authorship: Felipe Martins[1], Megan Phifer-Rixey[2], and Michael W. Nachman[3]

[1] *Research School of Biology, The Australian National University, Canberra, Australia.*
[2] *Department of Biology, Monmouth University, West Long Branch, NJ, 07764, USA.*
[3] *Department of Integrative Biology and Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, CA, 94720, USA.*

Abstract

Bergmann's rule is the observation that animals at more extreme latitudes have a larger body mass compared with animals at lower latitudes, presumably reflecting thermoregulatory adaptation. Recent studies have demonstrated that the gut microbiota can provide greater energy extraction from diet by breaking down indigestible polysaccharides and producing short-chain fatty acids (SCFAs). The extent to which gut microbiota plays a role in adaptive body mass variation remains largely unexplored. Here, we collected wild house mice from three latitudinal transects across the Americas and tested the association between gut microbial variation and host body mass variation. We found that overall differences in the gut microbial community were significantly associated with differences in host body mass accounting for geographic distance in two out of the three transects. To test whether the link between the gut microbiota and body mass are driven by other environmental factors, we conducted a common garden experiment by rearing offspring of wild-caught mice in captivity. We identified several gut microbial measurements that associated with body mass in wild-caught mice that persisted in lab-reared mice including; (1) phylogenetic diversity and (2) relative abundances of three bacterial genera that have previously been associated with functions related to energy harvest in mouse models and humans. Lastly, to test whether the gut microbiota is involved in energy extraction, we measured the amounts of fecal SCFAs among five wild-derived inbred mice from high and low latitudes that differ in body mass. We found mice from colder environments are larger and tend to produce greater amounts of SCFAs without an increase in food consumption compared to mice from warmer environments. Together, the results raise the intriguing possibility that the gut microbiota may be playing a role in the pattern described by Bergmann's rule, a fundamental pattern in evolutionary biology.

6.1. Introduction

A major goal in evolutionary biology is to link individual processes to macro-ecological and evolutionary patterns. Accumulating evidence suggests that the gut microbiota affects a variety of host phenotypes and thus an organisms's fitness (McFall-Ngai *et al.* 2013; Suzuki 2017). The beneficial role of the gut microbiota in host biology raises the interesting possibility that the gut microbiota may mediate environmental adaptations of their host. However, identifying the causal role of gut microbiotas in host adaptation remains a challenge.

Climatic adaptation has been proposed to explain the geographic variation of human gut microbiota where an obesity-associated gut microbial composition tend to be more prevalent in colder environments (Suzuki and Worobey 2014). Studies in obesity research suggest that the gut microbiota can play a causal role in host body mass variation at least under certain dietary conditions (Bäckhed *et al.* 2004; Ley *et al.* 2005; Turnbaugh *et al.* 2006). Increased energy extraction through bacteria-dependent digestion of plant polysaccharides, activation of fat storage, and production of short-chain fatty acids (SCFAs) has been proposed to explain the link between microbiota and body mass (Bäckhed *et al.* 2004; Turnbaugh *et al.* 2006). Obese individuals of mice and humans are characterized by the higher ratio of relative abundances of two dominant bacterial phyla, Firmicutes and Bacteroidetes (F/B ratio) (Ley *et al.* 2006). Germ-free mice that received the obesity-associated microbiota (higher F/B ratio) show significant increase in energy extraction and concentrations of cecal SCFAs compared to controls (Turnbaugh *et al.* 2006), a pattern also found in humans (Jumpertz *et al.* 2011). A significant positive correlation between F/B ratio and absolute latitude was found in a meta-analysis of human populations suggesting a possible role for the gut microbiota in mediating the differences in body mass (Suzuki and Worobey 2014). It is generally accepted that humans follow Bergmann's rule, the observation that animals living in higher latitudes have a larger body mass compared to animals living in lower latitudes, presumably reflecting thermoregulatory adaptation (Bergmann 1847; Roberts 1953).

Although the observation of human gut microbial patterns are consistent with the involvement of gut microbiota in climatic adaptation, factors affecting the gut microbiota and the causation of the microbiota-body mass relationship are unclear. For example, the F/B ratio is also associated with many other factors including diet (De Filippo *et al.* 2010), age (Mariat *et al.* 2009), and inflammation (Hansen, Gulati and Sartor 2010). Thus, manipulative experiments are necessary to further investigate the role of gut microbiota in potentially adaptive body mass variation. However, manipulative experiments in human subjects are extremely difficult if not impossible. Laboratory mouse models are often used as a stand in for human subjects to infer the host-microbial interactions, but the results may not translate between species (Nguyen *et al.* 2015). Therefore, merely describing the microbial patterns in the field or in manipulative experiments in model organisms may not be sufficient to identify the driver and function of the microbiota in natural populations. A system that can combine both approaches is necessary.

House mice (*Mus musculus*) provide a unique opportunity to study the role of gut microbiota in climatic adaptation by combining field observations and manipulative experiments in a single system. House mice have a global distribution in association with humans which encompasses a wide range of latitudes and climates (Guénet and Bonhomme 2003). Despite the recent colonization of house mice in the Americas (i.e.

probably in the past several hundred years), they show a clinal variation in body mass along the east coast of North America (Lynch 1992), a pattern expected by Bergmann's rule. The population differences in body mass persist in a common environment after multiple generations (Lynch 1992), suggesting that body mass variation may reflect a genetic adaptation in response to thermoregulatory stress. Experimental evidence demonstrates that wild house mice kept in cold temperatures increase body weight and fat compared to controls in fewer than 10 generations (Barnett and Dickson 1989). The gut microbiota of wild house mice is also known to differ by geographic and genetic distances (Linnenbrink *et al.* 2013) and the alpha-diversity of the gut microbiota has been associated with differences in body weight (Weldon *et al.* 2015). Wild house mice can be easily kept in captivity to test hypotheses on their gut microbiota by conducting experiments (Wang *et al.* 2014, 2015) and classic laboratory mouse strains have been studied extensively as a mammalian model in gut microbial ecology and obesity research.

Here, we tested associations between gut microbiota variation and body mass variation in wild house mice collected from three latitudinal transects across the Americas by combining field observations and laboratory experiments. First, we tested whether differences in body mass significantly correlate with beta-diversity of the gut microbiota in natural populations. Next, we conducted a common garden experiment in the laboratory to identify microbial measurements that consistently associate with body mass variation in both wild and lab environments. Finally, we quantified fecal SCFAs from five wild-derived inbred lines representing differences in body mass from a range of thermal environments. Overall, the results support a potential role of gut microbiota in thermoregulatory adaptation.

6.2. Methods

6.2.1. Animal collection

A total of 162 individuals of wild house mice were collected from three latitudinal transects across the Americas (Fig.1A). The East-NA transect includes the same individuals from the five populations described in Chapter 4 collected in summer 2012. The West-NA transect includes five populations collected in summer 2012; (1) Arizona, (2) Southern Utah, (3) Northern Utah, and (4) Montana in the U.S. and (5) Edmonton in Canada. The SA transect includes seven populations collected between February and September 2013; (1) Porto Velho, (2) Brasilia, (3) Maringa, and (4) Uruguaiana in Brazil and (5) Tandil, (6) Gaiman, and (7) Ushuaia in Argentina. Each individual was separated by a minimum of 500m to avoid collecting close relatives except two sites in Edmonton and two sites in Ushuaia. Detailed sample information is provided in Table S1. Body weight, body mass index (i.e. BMI = body weight / body length squared), and cecal samples were collected within 24 hours after capture. Cecal samples were flash-frozen in the field using liquid nitrogen and stored in -80°C until DNA extraction. Following the protocol of Suzuki and Nachman (2016), carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) stable isotopes from mouse hair was analyzed to infer diet. We used WorldClim database (Hijmans *et al.* 2005) to download 19 climatic variables based on GPS localities using the R package "dismo". Principle components were calculated in JMP 14.0 (SAS institute)

and the first two PC axes, climate PC1 (48.8%) and climate PC2 (21.8%) were used for the downstream analyses. All procedures involving animals were approved by the Institutional Animal Care and Use Committee at the University of Arizona (07-004) and the University of California Berkeley (R361-0514). Museum specimens (skins and skulls) and associated data have been deposited in the Museum of Vertebrate Zoology at the University of California, Berkeley and uploaded to a public database ARCTOS.

6.2.2. Lab-reared animals and wild-derived inbred lines

Live animals were collected from five locations across the Americas; (1) Manaus, Amazonas, Brazil (BR), (2) Gainsville, Florida, USA (FL), (3) Tucson, Arizona, USA (AZ), (4) Saratoga Springs, New York, USA (NY) and (5) Edmonton, Alberta, Canada (EDM) (Fig.1A). Wild-caught animals were transferred and maintained in a standard lab environment; 23°C with 10 hours dark and 14 hour light cycles and Teklad Global food (18% Protein Rodent Diet) was fed *ad libitum*. Sib-sib mating was utilized to create multiple wild-derived inbred lines from multiple wild-caught parents within each population. For the common garden experiment, we collected body weight data and fresh fecal samples from a total of 120 individuals representing 40 individuals each from three populations; BR, FL, and NY. Four adult individuals (i.e. two males and two females) from 10 independent crosses of wild-caught founders were used to represent each population (Table S2).

6.2.3. Quantification of fecal SCFAs and associated measurements

We measured the concentration of fecal SCFAs from all five wild-derived inbred lines mentioned above. Five to seven age-matched male individuals from later generation (i.e. up to generation 14) per population were used for the experiment (Table S3). Each individual male represent an independent inbred line (i.e. descendants of different wild-caught parents). Along with the collection of fresh fecal samples for SCFA quantification, body weight, weight of food intake, and weight of feces were measured in individuals between the age of 100-112 days. The amounts of food intake and defecation were measured every 24 hours for three consecutive days and each measurement was averaged. Both measurements were divided by body weight to account for body mass differences. The fresh fecal samples were immediately stored in -80°C. The frozen fecal samples were shipped to the West Coast Metabolomics Center at the University of California Davis for quantification of SCFAs, including acetate, propionate, and butyrate. Mass spectrometry was conducted on an Agilent 6890A Gas Chromatograph with an Agilent 5977A Mass Selective Detector. The method has been validated and explained in detail (Richardson 1989; Moreau *et al.* 2003).

6.2.4. DNA extraction and 16S amplicon sequencing

To characterize the gut microbiota, frozen cecal samples were used for the wild-caught individuals and frozen fecal samples were used for lab-reared individuals. Although the sample types are different, we know that fecal samples represent individual differences in the cecal microbial community at least in wild house mice (Suzuki and

Nachman 2016). We followed the protocol of Suzuki and Nachman (2016) for the DNA extraction and 16S rRNA sequencing. Briefly, we used the QIAamp DNA stool Minikit (Qiagen) with a modified protocol adding a bead-beating step. The V4 region of the 16S rRNA gene was sequenced with 150bp paired-end Illumina MiSeq at the Next Generation Sequencing Core Facility at Argonne National Laboratory. The PCR primers (515F and 806R) and the barcodes are described in Caporaso *et al.* (2012).

6.2.5. Data analyses

The 16S rRNA data were demultiplexed and quality-filtered using default parameters in Qiime version 1.9.0 (Caporaso *et al.* 2010). A subsampled open-reference OTU picking approach was used with default parameters generating 97% OTUs. Singletons were removed. Bacterial taxonomy was assigned based on the SILVA database (release 128) (Quast *et al.* 2013) and phylogenetic trees were computed using FastTree (Price, Dehal and Arkin 2009). All samples were rarefied to 5,000 reads. Eight samples were excluded from all the downstream analyses due to the low number of reads (< 2000); FMM111, FMM112, FMM148, FMM274, BR07F1, BR08M1, FL08M1, and FL08M2.

Beta-diversity of the gut microbiota was calculated using Bray-Curtis dissimilarity (BCD) and PCoA plots were generated. Mantel tests were used to test correlations between BCD and seven predictor variables (i.e. geographic distance (km), climate PC1, climate PC2, body weight (g), BMI, δ^{13}C diet, and δ^{15}N diet) using all individuals combined and each transect separately. Similarly, Partial Mantel test was used to control the effect of geographic distance and test the correlation between BCD and rest of the six predictor variables. A Spearman's rho correlation was used to test correlations between absolute latitude and metadata. Alpha-diversity was calculated using phylogenetic diversity (PD) (Faith 1992). Correlations between PD and body weight were calculated using a Spearman's rho correlation for all wild-caught individuals, all lab-reared individuals, and each transect separately. To identify bacterial genera that correlate with body mass variation, we looked for relative abundances of bacteria that show correlations with body weight in the same direction between wild-caught individuals and lab-reared individuals. We selected 25 bacterial genera in which the average relative abundances were greater than 0.1% across all samples and were present in both wild and lab samples. Correlations were based on Spearman's rho correlation. Body weight, amount of food intake, amount of feces defecated, and concentrations of SCFAs were compared among five wild-derived inbred lines in a common environment using Kruskal-Wallis test. Pairwise comparisons were preformed using Wilcoxon test.

6.3. Results and discussion

6.3.1. Geography and body mass are associated with the compositional variation in the gut microbiota

To identify factors that correlate with beta-diversity of the gut microbiota, we first tested the association between BCD and seven predictor variables using all individuals

from three latitudinal transects across the Americas (Fig.1A, Table 1). We found geographic distance (Fig. 1B, Mantel r = 0.41, p < 0.0001), BMI (Fig. 1C, Mantel r = 0.14, p = 0.002), and δ^{13}C diet (Mantel r = 0.08, p = 0.03) correlated with BCD. Climate PCs, body weight, and δ^{15}N diet did not show significant correlations with BCD. After controlling for multiple testing, only geographic distance and BMI remained significant.

The association between body mass and the gut microbiota is interesting because the body weight and BMI show positive correlations with absolute latitude using all individuals (Table S4, Body weight: rho = 0.27m p = 0.0007, BMI: rho = 0.31, p < 0.0001), a pattern consistent with Bergmann's rule. For example, positive correlations between body weight and absolute latitude were observed in both East-NA (rho = 0.32, p = 0.025) and SA transects (rho = 0.26, p = 0.045) independently, although not in West-NA transect (rho = -0.28, p = 0.06). A previous study has also observed a similar pattern among house mouse populations on the east coast of North America (Lynch 1992). However, the association between body mass and gut microbiota using all individuals is confounded by many factors. For example, body mass also differs between North and South America (Fig. 1B&C), where mice from North America are on average larger than mice from South America in terms of body weight (North mean ± SD = 16.1±3.9 and South mean ± SD = 13.5±3.7, Student's t-test p < 0.0001) and BMI (North mean ± SD = 0.0023±0.0003 and South mean ± SD = 0.0018±0.0003, Student's t-test p < 0.0001). Thus, the overall association between body mass and BCD can potentially be explained by any environmental differences between North and South America.

To test whether the link between BCD and body mass persists within each transect, we tested correlations between BCD and predictor variables in three transects independently (Table S5). We found the effects of geographic distance and body mass on BCD differs among transects. For example, within the East-NA transect, BMI showed the strongest correlation with BCD (Mantel r = 0.25, p = 0.002). Similarly, within the West-NA transect, body weight (Mantel r = 0.21, p = 0.008) and geographic distance (Mantel r = 0.14, p = 0.004) showed the strongest correlations with BCD while within the SA transect, body mass measurements did not correlate with BCD and geographic distance (Mantel r = 0.35, p < 0.001) and climate PC1 (Mantel r = 0.39, p < 0.001) showed the strongest correlations with BCD (Table S5).

Lastly, we ran Partial Mantel tests to ask whether the correlations between body mass and BCD remain significant after controlling for the effect of geographic distance within each transect (Table S6). The overall pattern remained the same. For example, only BMI (Partial Mantel r = 0.25, p < 0.001) and body weight (Partial Mantel r = 0.21, p = 0.005) remained significant after controlling for geographic distance within East-NA and West-NA transect, respectively. Similarly, only climate PC1 remained significant after controlling for geographic distance (Partial Mantel r = 0.18, p < 0.001) within SA transect. Together, the results suggest that differences in body mass measurements predict the differences in the gut microbial composition in the two North American transects, but not in the South American transect.

6.3.2. Alpha-diversity of the gut microbiota positively correlate with body weight in wild-caught and lab-reared mice.

To further understand the link between body mass and gut microbiota, we asked whether alpha-diversity of the gut microbiota explain the variation in body mass using PD (Faith 1992) as observed in a previous study of wild house mice (Weldon *et al.* 2015). Similar to the result of beta-diversity, significant correlations between body weight and PD were observed in the East-NA transect (Fig.2A, rho = 0.46, p = 0.0008) and the West-NA transect (Fig.2B rho = 0.52, p = 0.0002), but not in the SA transect (Fig.2C rho = -0.03, p = 0.83). Although the SA transect did not show a significant correlation between body weight and PD, the correlation was robust when all wild-caught individuals were plotted together (Fig.2D, rho = 0.30, p = 0.0001). The positive correlation between body mass and alpha-diversity has been observed within species (Weldon *et al.* 2015) and between species (Godon *et al.* 2016; Nishida and Ochman 2017) of mammals. However, the mechanism of this relationship is unknown. Geography, climate, and diet are all possible explanations for the link between body mass and alpha-diversity in our field-collected data because they are all correlated with alpha-diversity to a certain degree (Table S7).

To test whether the link between body mass and alpha-diversity is independent of other variables such as geography, climate, diet, and environmental microbes, we conducted a common garden experiment. We collected live mice from three locations (i.e. NY, FL, and BR) and generated 40 animals from each location by crossing wild-caught parents in a common lab environment (Fig.1A). We found population differences in body mass persisted in a common environment after one generation (Fig. S1) consistent with a previous study (Lynch 1992). Surprisingly, we found even a stronger positive correlation between body weight and PD in a common environment (Fig.2E, rho = 0.55, $p < 0.0001$) compared to the correlation observed in the field (Fig.2D). The results suggest that the association between body mass and alpha-diversity is independent of other covariates.

The link between body mass and alpha-diversity can be explained by two main hypotheses. The first hypothesis is where changes in the alpha-diversity is a cause of greater body mass. For example, stability-diversity relationship (Tilman and Downing 1994; Doak *et al.* 1998) has been used to explain the link between higher microbial diversity and greater need of fermentation for energy extraction (Lu *et al.* 2014; Suzuki and Nachman 2016) and thus affect the body mass. The second hypothesis is where the greater alpha-diversity in the gut microbiota is a consequence of larger body mass. For example, species-area relationship (Preston 1962) can explain the link where larger animals have greater "area" for microbes to colonize and thus the increase in rare OTUs may increase alpha-diversity measurements. Further investigation is necessary to understand the causality of the alpha-diversity and body mass relationship.

6.3.3. Identification of gut bacterial genera that are associated with body mass in wild-caught and lab-reared animals

To identify specific bacterial taxa that correlate with body mass, we tested correlations between body weight and relative abundances of bacterial genera (Fig.3). We identified a total of 25 bacterial genera that had relative abundances greater than 0.1% that are present in both the wild and lab individuals. Among the 25 bacterial genera, 14 of them showed correlations with body weight (Fig. 3A). A positive correlation between

body weight and *Mycoplasma* remained significant after multiple testing. Similarly, eight out of 25 bacterial genera showed correlations with body weight in lab-reared mice that were born in a common environment (Fig. 3B). *Prevotella*, unclassified genus of YS2, *Helicobacter* showed significant positive correlations with body weight and an unclassified genus of Rikenellaceae showed a significant negative correlation with body weight in lab-reared mice.

The relative abundances of three bacterial genera positively correlated with body weight in both wild-caught and lab-reared mice (Fig. 3). Interestingly, all three genera have been associated with energy extraction related functions. For example, *Prevotella* is known to breakdown "indigestible" polysaccharides and produce SCFAs (Kovatcheva-Datchary *et al.* 2015), which is an important energy source for the mammalian host including humans (Bergman 1990). Recently, *Prevotella*-dominated communities have been shown to produce greater amount of SCFAs (Chen *et al.* 2017). The association between *Helicobacter pylori* and body mass has been reported in many studies and actively debated. For example, two recent meta-analyses in humans show contrasting results where *H. pylori*-positive patients show greater BMI in one study (Upala *et al.* 2016) and *H. pylori* prevalence was inversely correlate with obesity prevalence in another (Lender *et al.* 2014). The endotoxin-producing family Desulfovibrionaceae was enriched in mouse models with impaired glucose tolerance and associated with calorie intake (Zhang *et al.* 2010). Together, links between the three genera and functions related to energy extraction have been reported at least in wild-caught mice, lab-reared mice, classic-mouse models, and humans.

6.3.4. Larger mice living in colder environments produce greater amounts of SCFAs

The hypothesis that greater energy is extracted by the gut microbiota at higher-latitudes has been proposed based on human populations where obesity-associated gut microbiota (i.e. higher F/B ratio) significantly increased with absolute latitude (Suzuki and Worobey 2014). Interestingly, mice from the SA-transect showed a consistent pattern with a significant positive correlation between F/B ratio and absolute latitude (rho = 0.26, p = 0.04, Fig. S2). In genetically obese mouse models, greater F/B ratio has been associated with increased SCFAs and energy extraction (Turnbaugh *et al.* 2006). Although the other two transects in wild mice (i.e. East-NA and West-NA) did not show significant correlations, a similar trend of greater F/B ratio in animals living in colder environments observed in humans and mice suggest the involvement of microbiota-produced energy source (e.g. SCFAs) in host body mass variation.

To test whether SCFAs are associated with adaptive differences in body mass, we measured amounts of fecal SCFAs in wild-derived inbred lines collected from five populations that vary in body mass; BR, AZ, FL, NY, and EDM (Fig. 1A&4A). We used 5-7 male individuals per population and measured body weight, amount of food intake, amount of defecation, and amount of SCFAs at age between 100-112 days in a common environment (Table S3). First, body weight showed clinal variation where body weight tends to increase from low latitude populations to high latitude populations, a pattern consistent with Bergmann's rule (Fig. 4A). The amounts of average food intake and defecation corrected for body weight did not differ among populations, but they showed a trend of negative correlations with latitude (Fig. 4B&C). Despite the trend of less food

intake and defecation in populations from colder environments, the total amount of fecal SCFAs was greatest in the two northern most populations among the five populations tested (Fig. 4D). This trend is consistent among all three major SCFAs produced by bacteria; acetate (Fig.4E), propionate (Fig.4F), and butyrate (Fig.4G). The greater amount of total SCFAs, acetate, and propionate observed in NY mice compared to FL mice is particularly interesting because the two populations are from the same East-NA transect that show clinal variation in body mass (Table S4) and alpha- and beta-diversity of the gut microbiota were best explained by body mass variation (Table S6&7). Furthermore, several host genes have been associated with the variation in the gut microbiota observed in the East-NA transect using the same individuals (Chapter 4).
The results suggest that the gut microbiota of larger mice living in colder environments produce greater amounts of SCFAs without a relative increase in food consumption. Obesity and overweight individuals has been associated with the increased amounts of cecal and fecal SCFAs in mice (Turnbaugh *et al.* 2006; Murphy *et al.* 2010) and humans (Schwiertz *et al.* 2010; Fernandes *et al.* 2014; Rahat-Rozenbloom *et al.* 2014). Surgically-treated obese individuals show decrease in the concentrations of fecal SCFAs (Ppatil *et al.* 2012). These observations suggest that the gut microbiota in larger mice may have increased capacity to extract energy from diet. Together, the results support the possibility that the gut microbiota may play a role in adaptive body mass difference in natural populations of mammals.

6.3.5. Conclusions

Overall, we found significant associations between gut microbiota variation and host body mass variation in wild house mice based on wild population samples and laboratory experiments. We identified candidate microbial measurements that associated with host body mass variation including alpha-diversity and three bacterial genera that show consistent associations in both field- and laboratory-environments. We also found greater concentrations of fecal SCFAs in larger mice living in higher-latitudes compared to smaller mice living in lower-latitudes without an increase in food consumption. The results suggest a beneficial function of the gut microbiota in colder environments by providing greater energy extraction. Together, the study suggests a potential role of the gut microbiota in thermoregulatory adaptation.

6.4. Chapter 6 Tables

Table 1. Correlations between predictor variables and Bray-Curtis dissimilarity using Mantel test.

Predictor variables	All transects		
	n	Mantel r	p-value
Geographic distance	162	**0.407**	**<0.0001**
Climate PC1	162	0.059	0.103
Climate PC2	162	-0.009	0.734
Body weight	159	0.071	0.117
BMI	159	**0.144**	**0.0016**
Carbon	156	0.083	0.034
Nitrogen	156	0.038	0.462

Significant correlations after Bonferroni correction within each transect are in bold ($p = 0.05/7 = 0.007$)

Table S2. Sample information of lab-reared individuals (n=120) (continued).

SampleID	Colony	Sex	Weight	Carbon	Nitrogen
NY01F1	New_York_Lab	F	17.8	-19.05	6.38
NY01F2	New_York_Lab	F	17.9	-19.05	6.38
NY01M1	New_York_Lab	M	20.1	-19.05	6.38
NY01M2	New_York_Lab	M	21.5	-19.05	6.38
NY02F1	New_York_Lab	F	13.5	-19.05	6.38
NY02F2	New_York_Lab	F	16.8	-19.05	6.38
NY02M1	New_York_Lab	M	21.0	-19.05	6.38
NY02M2	New_York_Lab	M	24.8	-19.05	6.38
NY03F1	New_York_Lab	F	17.9	-19.05	6.38
NY03F2	New_York_Lab	F	22.1	-19.05	6.38
NY03M1	New_York_Lab	M	20.1	-19.05	6.38
NY03M2	New_York_Lab	M	22.0	-19.05	6.38
NY04F1	New_York_Lab	F	16.4	-19.05	6.38
NY04F2	New_York_Lab	F	17.8	-19.05	6.38
NY04M1	New_York_Lab	M	22.6	-19.05	6.38
NY04M2	New_York_Lab	M	23.9	-19.05	6.38
NY05F1	New_York_Lab	F	17.5	-19.05	6.38
NY05M1	New_York_Lab	M	23.4	-19.05	6.38
NY05M2	New_York_Lab	M	25.1	-19.05	6.38
NY05M3	New_York_Lab	M	21.9	-19.05	6.38
NY06F1	New_York_Lab	F	15.0	-19.05	6.38
NY06F2	New_York_Lab	F	16.5	-19.05	6.38
NY06M1	New_York_Lab	M	24.2	-19.05	6.38
NY06M2	New_York_Lab	M	25.2	-19.05	6.38
NY07F1	New_York_Lab	F	20.1	-19.05	6.38
NY07F2	New_York_Lab	F	20.0	-19.05	6.38
NY07M1	New_York_Lab	M	20.9	-19.05	6.38
NY07M2	New_York_Lab	M	24.8	-19.05	6.38
NY08F1	New_York_Lab	F	13.8	-19.05	6.38
NY08F2	New_York_Lab	F	15.0	-19.05	6.38
NY08M1	New_York_Lab	M	17.0	-19.05	6.38
NY08M2	New_York_Lab	M	16.2	-19.05	6.38
NY09F1	New_York_Lab	F	21.1	-19.05	6.38
NY09F2	New_York_Lab	F	23.2	-19.05	6.38
NY09M1	New_York_Lab	M	24.5	-19.05	6.38
NY09M2	New_York_Lab	M	27.0	-19.05	6.38
NY10F1	New_York_Lab	F	17.0	-19.05	6.38
NY10F2	New_York_Lab	F	20.9	-19.05	6.38
NY10M1	New_York_Lab	M	24.2	-19.05	6.38
NY10M2	New_York_Lab	M	25.5	-19.05	6.38
BR01F1	Brazil_Lab	F	14.7	-18.91	6.20
BR01F2	Brazil_Lab	F	15.2	-18.91	6.20
BR01M1	Brazil_Lab	M	16.5	-18.91	6.20
BR01M2	Brazil_Lab	M	15.5	-18.91	6.20
BR02F1	Brazil_Lab	F	13.5	-18.91	6.20
BR02F2	Brazil_Lab	F	10.1	-18.91	6.20
BR02F3	Brazil_Lab	F	13.2	-18.91	6.20
BR02M1	Brazil_Lab	M	15.8	-18.91	6.20
BR03F1	Brazil_Lab	F	12.6	-18.91	6.20
BR03F2	Brazil_Lab	F	12.4	-18.91	6.20
BR03M1	Brazil_Lab	M	16.1	-18.91	6.20
BR03M2	Brazil_Lab	M	17.9	-18.91	6.20
BR04F1	Brazil_Lab	F	9.5	-18.91	6.20
BR04F2	Brazil_Lab	F	15.2	-18.91	6.20
BR04M1	Brazil_Lab	M	15.5	-18.91	6.20
BR04M2	Brazil_Lab	M	16.0	-18.91	6.20
BR05F1	Brazil_Lab	F	10.0	-18.91	6.20
BR05F2	Brazil_Lab	F	12.9	-18.91	6.20
BR05F3	Brazil_Lab	F	12.2	-18.91	6.20
BR05M1	Brazil_Lab	M	12.5	-18.91	6.20
BR06F1	Brazil_Lab	F	12.0	-18.91	6.20

Table S2. Continued.

SampleID	Colony	Sex	Weight	Carbon	Nitrogen
BR06F2	Brazil_Lab	F	12.9	-18.91	6.20
BR06M1	Brazil_Lab	M	15.5	-18.91	6.20
BR06M2	Brazil_Lab	M	15.1	-18.91	6.20
BR07F1	Brazil_Lab	F	13.1	-18.91	6.20
BR07F2	Brazil_Lab	F	10.5	-18.91	6.20
BR07F3	Brazil_Lab	F	10.1	-18.91	6.20
BR07M1	Brazil_Lab	M	14.2	-18.91	6.20
BR08F1	Brazil_Lab	F	12.1	-18.91	6.20
BR08F2	Brazil_Lab	F	13.3	-18.91	6.20
BR08M1	Brazil_Lab	M	17.5	-18.91	6.20
BR08M2	Brazil_Lab	M	12.5	-18.91	6.20
BR09F1	Brazil_Lab	F	12.6	-18.91	6.20
BR09F2	Brazil_Lab	F	13.0	-18.91	6.20
BR09F3	Brazil_Lab	F	14.0	-18.91	6.20
BR09M1	Brazil_Lab	M	16.4	-18.91	6.20
BR10F1	Brazil_Lab	F	11.8	-18.91	6.20
BR10F2	Brazil_Lab	F	9.5	-18.91	6.20
BR10M1	Brazil_Lab	M	15.6	-18.91	6.20
BR10M2	Brazil_Lab	M	15.0	-18.91	6.20
FL01F1	Florida_Lab	F	16.6	-18.98	6.68
FL01F2	Florida_Lab	F	19.0	-18.98	6.68
FL01M1	Florida_Lab	M	23.6	-18.98	6.68
FL01M2	Florida_Lab	M	21.0	-18.98	6.68
FL02F1	Florida_Lab	F	15.7	-18.98	6.68
FL02F2	Florida_Lab	F	16.2	-18.98	6.68
FL02M1	Florida_Lab	M	20.2	-18.98	6.68
FL02M2	Florida_Lab	M	23.5	-18.98	6.68
FL03F1	Florida_Lab	F	17.5	-18.98	6.68
FL03F2	Florida_Lab	F	20.5	-18.98	6.68
FL03M1	Florida_Lab	M	25.0	-18.98	6.68
FL03M2	Florida_Lab	M	27.2	-18.98	6.68
FL04F1	Florida_Lab	F	12.9	-18.98	6.68
FL04F2	Florida_Lab	F	15.5	-18.98	6.68
FL04M1	Florida_Lab	M	20.0	-18.98	6.68
FL04M2	Florida_Lab	M	20.9	-18.98	6.68
FL05F1	Florida_Lab	F	11.9	-18.98	6.68
FL05F2	Florida_Lab	F	11.8	-18.98	6.68
FL05M1	Florida_Lab	M	19.5	-18.98	6.68
FL05M2	Florida_Lab	M	15.8	-18.98	6.68
FL06F1	Florida_Lab	F	16.1	-18.98	6.68
FL06F2	Florida_Lab	F	12.5	-18.98	6.68
FL06M1	Florida_Lab	M	19.2	-18.98	6.68
FL06M2	Florida_Lab	M	14.2	-18.98	6.68
FL07F1	Florida_Lab	F	13.5	-18.98	6.68
FL07F2	Florida_Lab	F	15.6	-18.98	6.68
FL07M1	Florida_Lab	M	15.4	-18.98	6.68
FL07M2	Florida_Lab	M	16.9	-18.98	6.68
FL08F1	Florida_Lab	F	18.5	-18.98	6.68
FL08F2	Florida_Lab	F	22.0	-18.98	6.68
FL08M1	Florida_Lab	M	21.1	-18.98	6.68
FL08M2	Florida_Lab	M	21.5	-18.98	6.68
FL09F1	Florida_Lab	F	15.0	-18.98	6.68
FL09F2	Florida_Lab	F	13.0	-18.98	6.68
FL09M1	Florida_Lab	M	17.5	-18.98	6.68
FL09M2	Florida_Lab	M	20.1	-18.98	6.68
FL10F1	Florida_Lab	F	15.5	-18.98	6.68
FL10F2	Florida_Lab	F	15.0	-18.98	6.68
FL10M1	Florida_Lab	M	18.8	-18.98	6.68
FL10M2	Florida_Lab	M	18.4	-18.98	6.68

Table S3. Sampled picked for SCFA and energy intake, age 100 - 112 day old animals.

			Age (days)	
Colony	Sex	n	Mean	SD
Edmonton	M	5	106.6	3.8
New York	M	7	106.4	3.9
Florida	M	7	106.9	4.1
Arizona	M	6	105.0	4.3
Brazil	M	7	106.6	3.3

Table S4. Correlations with absolute latitude and metadata

Variables	All samples		East-NA		West-NA		SA	
	rho	p-value	rho	p-value	rho	p-value	rho	p-value
Climate PC1	**-0.91**	**<.0001**	**-0.97**	**<.0001**	**-0.87**	**<.0001**	**-0.98**	**<.0001**
Climate PC2	0.08	0.30	0.32	0.022	**0.85**	**<.0001**	0.12	0.34
Body weight	**0.27**	**0.0007**	0.32	0.025	-0.28	0.061	0.26	0.045
Body mass index (BMI)	**0.31**	**<.0001**	0.21	0.14	-0.17	0.26	0.18	0.15
Diet (Carbon)	-0.18	0.021	0.19	0.19	-0.37	0.0088	-0.21	0.11
Diet (Nitrogen)	-0.09	0.27	0.14	0.35	0.00	0.98	-0.01	0.95

Significant correlations after Bonferroni correction within each transect are bolded (p = 0.05/6 = 0.008)

Table S5. Correlations between predictor variables and Bray-Curtis dissimilarity using Mantel test.

Predictor variables	All			East-NA			West-NA			SA		
	n	Mantel r	p-value	n	Mantel r	p-value	n	Mantel r	p-value	n	Mantel r	p-value
Geographic distance	162	**0.407**	**0.0001**	50	0.074	0.081	50	**0.135**	**0.004**	62	**0.351**	**<0.001**
Climate PC1	162	0.059	0.103	50	0.104	0.066	47	0.146	0.017	62	**0.388**	**<0.001**
Climate PC2	162	-0.009	0.734	50	0.132	0.039	47	0.116	0.013	62	0.031	0.47
Body weight	159	0.071	0.117	50	0.098	0.132	47	0.206	0.008	62	0.090	0.226
BMI	159	**0.144**	**0.0016**	50	**0.248**	**0.002**	47	0.123	0.151	62	0.064	0.392
Carbon	156	0.083	0.034	49	0.138	0.047	45	0.126	0.051	62	-0.012	0.812
Nitrogen	156	0.038	0.462	49	0.007	0.927	45	0.048	0.665	62	-0.027	0.684

Significant correlations after Bonferroni correction within each transect are bolded (p = 0.05/7 = 0.007)

Table S6. Correlations between predictor variables and Bray-Curtis dissimilarity using Partial Mantel test controlling for geographic distance.

Predictor variables	All			East-NA			West-NA			SA		
	n	Mantel r	p-value	n	Mantel r	p-value	n	Mantel r	p-value	n	Mantel r	p-value
Climate PC1	162	-0.057	0.933	50	0.100	0.07	47	0.051	0.242	62	**0.180**	**<0.001**
Climate PC2	162	-0.056	0.979	50	0.111	0.041	47	0.024	0.371	62	-0.004	0.539
Body weight	159	0.046	0.153	50	0.091	0.074	47	**0.211**	**0.005**	62	0.091	0.108
BMI	159	0.037	0.201	50	**0.245**	**<0.001**	47	0.124	0.092	62	0.062	0.203
Carbon	156	0.075	0.036	49	0.142	0.021	45	0.122	0.035	62	-0.010	0.532
Nitrogen	156	0.011	0.378	49	0.000	0.494	45	0.060	0.26	62	0.005	0.41

Significant correlations after Bonferroni correction within each transect are bolded (p = 0.05/7 = 0.007)

Table S7. Correlations between phylogenetic diversity (alpha-diversity) and metadata

Variables	All samples		East-NA		West-NA		SA	
	rho	p-value	rho	p-value	rho	p-value	rho	p-value
Latitude (absolute)	-0.1666	0.0341	0.3323	0.0184	-0.1075	0.4573	**-0.6014**	**<.0001**
Latitude	**0.2659**	**0.0006**	0.3323	0.0184	-0.1075	0.4573	**0.6014**	**<.0001**
Climate PC1	0.0391	0.6214	**-0.3721**	**0.0078**	0.015	0.9177	**0.6022**	**<.0001**
Climate PC2	0.0284	0.7202	-0.1515	0.2935	0.095	0.5117	0.1631	0.2052
Body weight	**0.3024**	**0.0001**	**0.4584**	**0.0008**	**0.5156**	**0.0002**	-0.0271	0.8345
Body mass index (BMI)	**0.2549**	**0.0012**	0.3329	0.0182	**0.4332**	**0.0024**	-0.0262	0.84
Diet (Carbon)	0.1724	0.0298	0.187	0.1983	**0.4106**	**0.0037**	0.131	0.31
Diet (Nitrogen)	-0.0317	0.6917	-0.0231	0.8749	-0.0341	0.818	0.0636	0.6233

Significant correlations after Bonferroni correction within each transect in bold (p = 0.05/6 = 0.008)

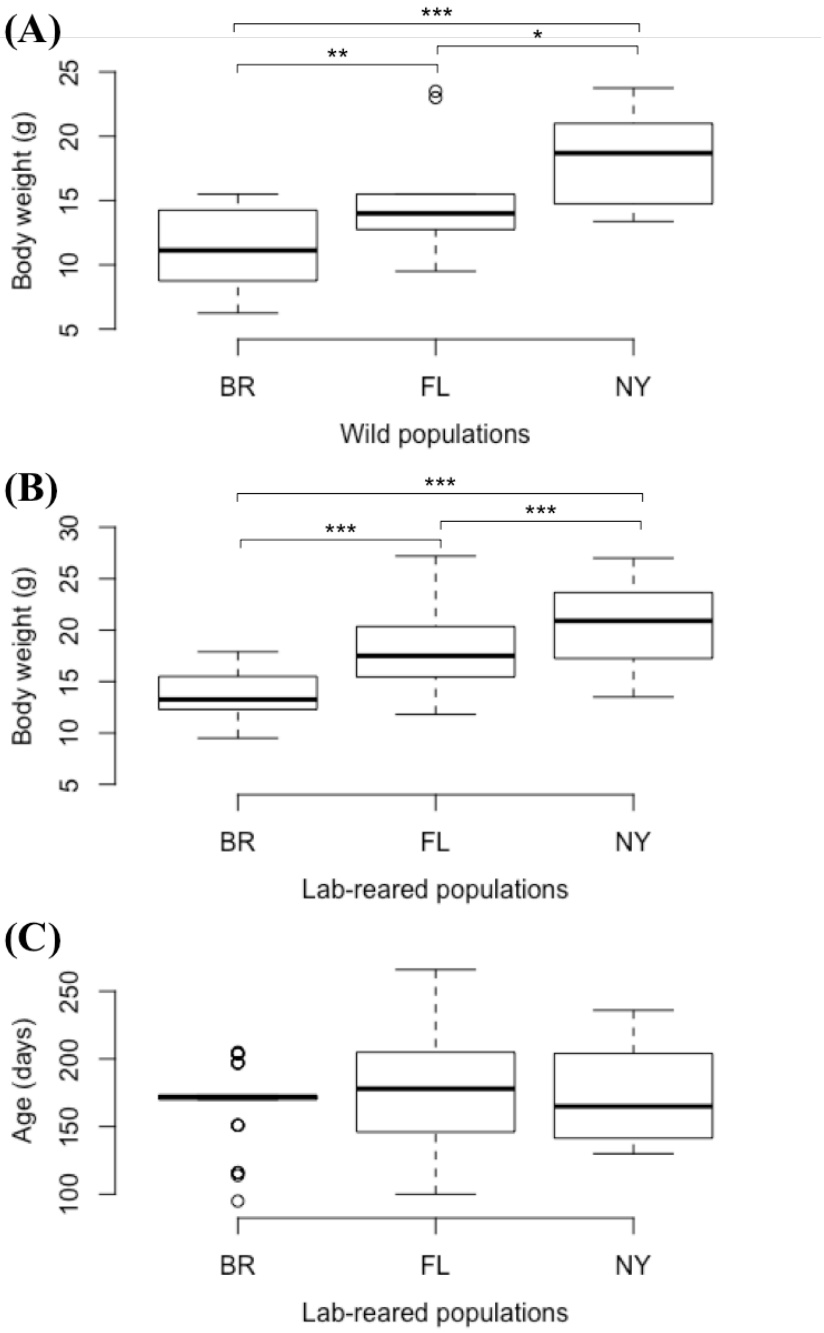


Figure S1. Population difference in body weight persists in a common environment. (A) Body weight differ among the wild populations (ANOVA $p = 0.0015$). (B) Body weight differ among lab-reared populations (ANOVA $p < 0.0001$). (C) Age does not differ among lab-reared populations (ANOVA $p = 0.33$). Pairwise significance is based on Student's t-test: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

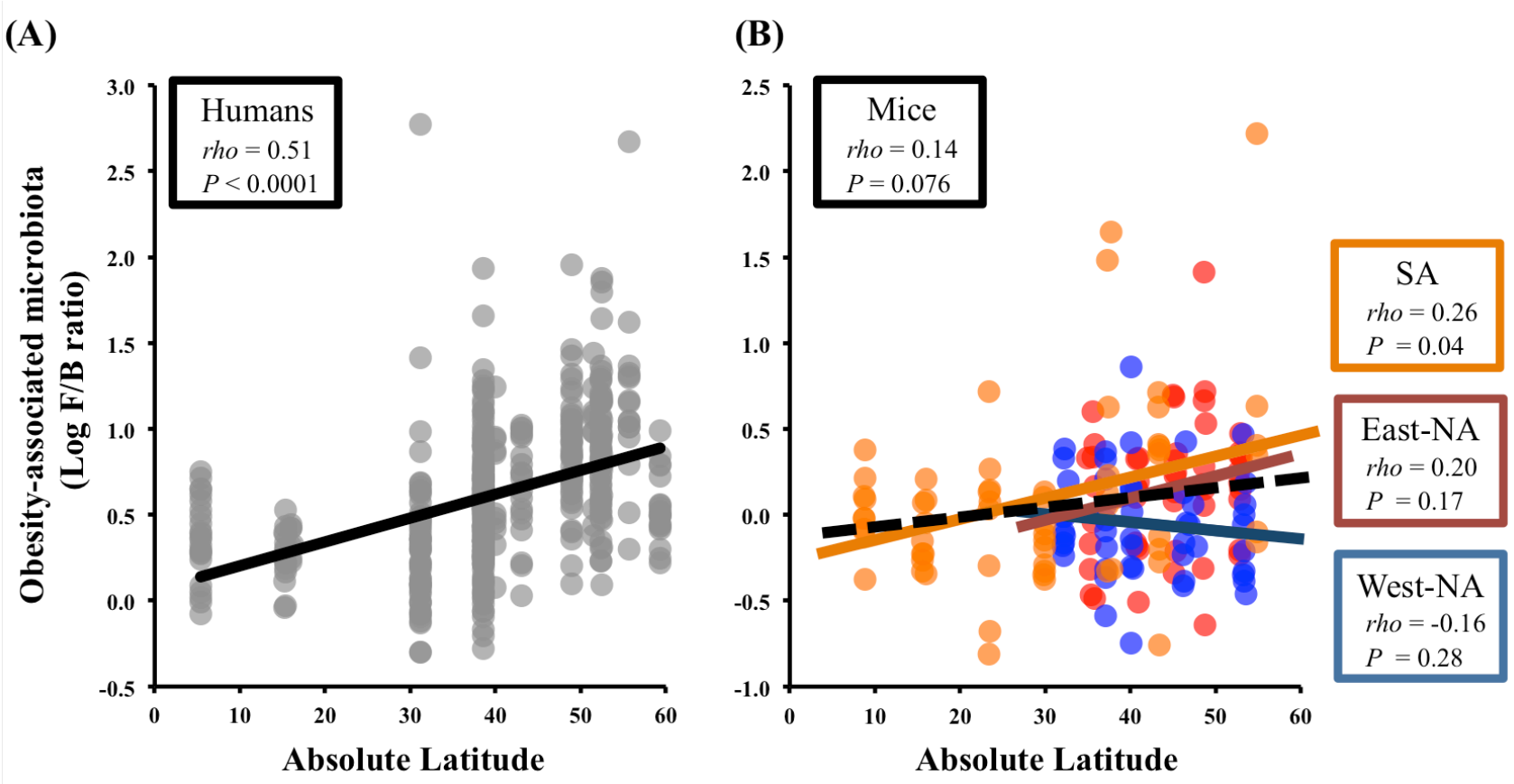


Figure S2. Correlations between log ratio of relative abundance of Firmicutes and Bacteroidetes (F/B ratio) and absolute latitude in humans and house mice. (A) In healthy adult humans, a significant positive correlation between log F/B ratio and absolute latitude was observed (rho = 0.51, p < 0.0001). (B) In wild house mice, positive slopes were observed between log F/B ratio and absolute latitude using all samples, SA transect, East-NA transect, but not West-NA transect. Only SA-transect showed a significant positive correlation using Spearman's rho correlation (rho = 0.26, p = 0.04).

Reference

Albenberg L, Esipova T V, Judge CP *et al.* Correlation Between Intraluminal Oxygen Gradient and Radial Partitioning of Intestinal Microbiota in Humans and Mice. *Gastroenterology* 2014;**147**:1055–1063.e8.

Albone ES. *Mammalian Semiochemistry, The Investigation of Chemical Signals Between Mammals*. John Wiley & Sons, Ltd., 1984.

Albone ES, Eglinton G, Walker JM *et al.* The anal sac secretion of the red fox (Vulpes vulpes); its chemistry and microbiology. A comparison with the anal sac secretion of the lion (Panthera leo). *Life Sci* 1974;**14**:387–400.

Albone ES, Perry GC. Anal sac secretion of the red fox, Vulpes vulpes; volatile fatty acids and diamines: Implications for a fermentation hypothesis of chemical recognition. *J Chem Ecol* 1976;**2**:101–11.

Alexy KJ, Gassett JW, Osborn DA *et al.* Bacterial fauna of the tarsal tufts of white-tailed deer (Odocoileus virginianus). *Am Midl Nat* 2003;**149**:237–40.

Allison A. Polymorphism and Natural Selection in Human Populations. *Cold Spring Harb Symp Quant Biol* 1964;**29**:137–49.

Amato KR, Leigh SR, Kent A *et al.* The Gut Microbiota Appears to Compensate for Seasonal Diet Variation in the Wild Black Howler Monkey (Alouatta pigra). *Microb Ecol* 2014;**69**:434–43.

Amato KR, Yeoman CJ, Kent A *et al.* Habitat degradation impacts black howler monkey (Alouatta pigra) gastrointestinal microbiomes. *ISME J* 2013:1–10.

Angelakis E, Yasir M, Bachar D *et al.* Gut microbiome and dietary patterns in different Saudi populations and monkeys. *Sci Rep* 2016;**6**:32191.

Archie EA, Theis KR. Animal behaviour meets microbial ecology. *Anim Behav* 2011;**82**:425–36.

Archie EA, Tung J. Social behavior and the microbiome. *Curr Opin Behav Sci* 2015;**6**:28–34.

Arumugam M, Raes J, Pelletier E *et al.* Enterotypes of the human gut microbiome. *Nature* 2011;**473**:174–80.

Bäckhed F, Ding H, Wang T *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A* 2004;**101**:15718–23.

Barcroft BYJ, Mcanally RA, Phillipson AT. Absorption of volatile acids from the alimentary tract of the sheep and other animals. *J Exp Biol* 1944;**20**:120–9.

Barnett S a, Dickson RG. Wild mice in the cold: some findings on adaptation. *Biol Rev Camb Philos Soc* 1989;**64**:317–40.
Barr JJ, Auro R, Furlan M *et al.* Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci U S A* 2013;**110**:10771–6.
Baumann P. Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. *Annu Rev Microbiol* 2005;**59**:155–89.
Benson AK, Kelly S a, Legge R *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* 2010a;**107**:18933–8.
Benson AK, Kelly S a, Legge R *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* 2010b;**107**:18933–8.
Bergman EN. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol Rev* 1990;**70**:567–90.
Bergmann C. Uber die verhaltnisse der warmeokonomie der thiere zu ihrer grosse. *Göttinger Stud Pt* 1847;**1**:595–708.
den Besten G, van Eunen K, Groen AK *et al.* The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J Lipid Res* 2013;**54**:2325–40.
Biddle A, Stewart L, Blanchard J *et al.* Untangling the genetic basis of fibrolytic specialization by lachnospiraceae and ruminococcaceae in diverse gut communities. *Diversity* 2013;**5**:627–40.
Blekhman R, Goodrich JK, Huang K *et al.* Host genetic variation impacts microbiome composition across human body sites. *Genome Biol* 2015;**16**:191.
Bonder MJ, Kurilshikov A, Tigchelaar EF *et al.* The effect of host genetics on the gut microbiome. *Nat Genet* 2016;**48**:1407–12.
Bordenstein SR, Theis KR. Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biol* 2015;**13**:1–23.
Brobeck J. Food intake as a mechanism of temperature regulation. *Yale J Biol Med* 1948;**20**:641–5.
Brooks AW, Kohl KD, Brucker RM *et al.* Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLoS Biol* 2016;**14**:e2000225.
Byndloss MX, Olsan EE, Rivera-Chávez F *et al.* Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science (80-)* 2017;**357**:570–5.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335–6.
Caporaso JG, Lauber CL, Walters W a *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 2012;**6**:1621–4.
Carey H V, Andrews MT, Martin SL. Mammalian hibernation: cellular and molecular responses to depressed metabolism and low temperature. *Physiol Rev* 2003;**83**:1153–81.
Carey H V, Walters W a, Knight R. Seasonal restructuring of the ground squirrel gut microbiota over the annual hibernation cycle. *Am J Physiol Regul Integr Comp Physiol* 2013;**304**:R33-42.
Carmody RN, Gerber GK, Luevano JM *et al.* Diet Dominates Host Genotype in Shaping the Murine Gut Microbiota. *Cell Host Microbe* 2015;**17**:72–84.
Carrillo-Araujo M, Tas N, Alcántara-Hernández RJ *et al.* Phyllostomid bat microbiome composition is associated to host phylogeny and feeding strategies. *Front Microbiol* 2015;**6**:1–9.
Chandler JA, Morgan Lang J, Bhatnagar S *et al.* Bacterial Communities of Diverse Drosophila Species: Ecological Context of a Host–Microbe Model System. Malik HS (ed.). *PLoS Genet* 2011;**7**:e1002272.
Chen T, Long W, Zhang C *et al.* Fiber-utilizing capacity varies in Prevotella- versus Bacteroides-dominated gut microbiota. *Sci Rep* 2017;**7**:1–7.
Chevalier C, Stojanovic O, Colin DJ *et al.* Gut microbiota orchestrates energy homeostasis during cold. *Cell* 2015;**163**:1360–74.
Choo JM, Leong LEX, Rogers GB. Sample storage conditions significantly influence faecal microbiome profiles. *Sci Rep* 2015;**5**:1–10.
Colombo BM, Scalvenzi T, Benlamara S *et al.* Microbiota and mucosal immunity in amphibians. *Front Immunol* 2015;**6**:1–15.
Comuzzie AG, Cole SA, Laston SL *et al.* Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population. *PLoS One* 2012;**7**, DOI: 10.1371/journal.pone.0051954.
Costello EK, Lauber CL, Hamady M *et al.* Bacterial community variation in human body habitats across space and time. *Science* 2009;**326**:1694–7.
Dai X, Zhang B, Wu X *et al.* Identification of tannin-degrading microorganisms in the gut of plateau pikas (Ochotona curzoniae) and root voles (Microtus oeconomus). *Symbiosis* 2014;**63**:1–9.
Davenport ER, Cusanovich DA, Michelini K *et al.* Genome-Wide Association Studies of the Human Gut Microbiota. 2015:1–22.

Delsuc F, Metcalf JL, Wegener Parfrey L *et al.* Convergence of gut microbiomes in myrmecophagous mammals. *Mol Ecol* 2014;**23**:1301–17.
Deng X, Sabino EC, Cunha-Neto E *et al.* Genome wide association study (GWAS) of chagas cardiomyopathy in trypanosoma cruzi seropositive subjects. *PLoS One* 2013;**8**:4–10.
Derrien M, Collado MC, Ben-Amor K *et al.* The mucin degrader Akkermansia muciniphila is an abundant resident of the human intestinal tract. *Appl Environ Microbiol* 2008;**74**:1646–8.
DeSantis TZ, Hugenholtz P, Larsen N *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;**72**:5069–72.
Dill-Mcfarland KA, Neil KL, Zeng A *et al.* Hibernation alters the diversity and composition of mucosa-associated bacteria while enhancing antimicrobial defence in the gut of 13-lined ground squirrels. *Mol Ecol* 2014;**23**:4658–69.
Doak DF, Bigger D, Harding EK *et al.* The statistical inevitability of stability-diversity relationships in community ecology. *Am Nat* 1998;**151**:264–76.
Dunbar HE, Wilson ACC, Ferguson NR *et al.* Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol* 2007;**5**:1006–15.
Durbán A, Abellán JJ, Jiménez-Hernández N *et al.* Assessing Gut Microbial Diversity from Feces and Rectal Mucosa. *Microb Ecol* 2011;**61**:123–33.
Eckburg PB, Bik EM, Bernstein CN *et al.* Diversity of the human intestinal microbial flora. *Science* 2005;**308**:1635–8.
Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;**26**:2460–1.
Edgar RC, Haas BJ, Clemente JC *et al.* UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011;**27**:2194–200.
Elsden SR, Hitchcock MWS, Marshall RA *et al.* olatie acid in the digesta of ruminants and other animals. *J Exp Biol* 1946;**22**:191–202.
Ephraim E, Odenyo A, Ashenafi M. Isolation and characterization of tannin-degrading bacteria from faecal samples of some wild ruminants in Ethiopia. *Anim Feed Sci Technol* 2005;**118**:243–53.
Espey MG. Role of oxygen gradients in shaping redox relationships between the human intestine and its microbiota. *Free Radic Biol Med* 2013;**55**:130–40.
Ezenwa VO, Williams AE. Microbes and animal olfactory communication: Where do we go from here? *BioEssays* 2014;**36**:847–54.

Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv* 1992;**61**:1–10.
Fang W, Fang Z, Zhou P *et al.* Evidence for Lignin Oxidation by the Giant Panda Fecal Microbiome. *PLoS One* 2012;**7**:1–10.
Feldhamer GA, Dickamer LC, Vessey SH *et al. Mammalogy: Adaptation, Diversity, Ecology*. 3rd ed. Baltimore: The Johns Hopkins University Press, 2007.
Fernandes J, Su W, Rahat-Rozenbloom S *et al.* Adiposity, gut microbiota and faecal short chain fatty acids are linked in adult humans. *Nutr Diabetes* 2014;**4**, DOI: 10.1038/nutd.2014.23.
De Filippo C, Cavalieri D, Di Paola M *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A* 2010;**107**:14691–6.
Fogel AT. The gut microbiome of wild lemurs: A comparison of sympatric lemur catta and propithecus verreauxi. *Folia Primatol* 2015;**86**:85–95.
Fraune S, Bosch TCG. Why bacteria matter in animal development and evolution. *BioEssays* 2010;**32**:571–80.
Gassett JW, Dasher KA, Miller KV *et al.* White-tailed deer tarsal glands : sex and age-related variation in microbial flora. *Mammalia* 2000;**64**, DOI: 10.1515/mamm.2000.64.3.371.
Godon JJ, Arulazhagan P, Steyer JP *et al.* Vertebrate bacterial gut diversity: Size also matters. *BMC Ecol* 2016;**16**:1–9.
Gomez A, Rothman JM, Petrzelkova K *et al.* Temporal variation selects for diet–microbe co-metabolic traits in the gut of Gorilla spp. *ISME J* 2015;**10**:514–26.
Goodrich JK, Davenport ER, Beaumont M *et al.* Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host Microbe* 2016a;**19**:731–43.
Goodrich JK, Davenport ER, Waters JL *et al.* Cross-species comparisons of host genetic associations with the microbiome. *Science (80-)* 2016b;**352**:29–32.
Goodrich JK, Waters JL, Poole AC *et al.* Human genetics shape the gut microbiome. *Cell* 2014;**159**:789–99.
Gorman ML. A mechanism for individual recognition by odour in Herpestes auropunctatus (Carnivora: Viverridae). *Anim Behav* 1976;**24**:141–5.
Gregory J., Nakano H. Pyridoxine and pyridoxine-5′-β-D-glucoside exert different effects on tissue B-6 vitamers but similar effects on β-glucosidase activity in rats. *J Nutr* 1995;**6**:2751–62.
Grindlay L, Regensteiner JG. Adaptation to high altitude. *Annu Rev Anthropol* 1983;**12**:285–304.
Gu S, Chen D, Zhang J-N *et al.* Bacterial community mapping of the mouse

gastrointestinal tract. *PLoS One* 2013;**8**:e74957.
Guénet JL, Bonhomme F. Wild mice: An ever-increasing contribution to a popular mammalian model. *Trends Genet* 2003;**19**:24–31.
Hall AB, Tolonen AC, Xavier RJ. Human genetic variation and the gut microbiome in disease. *Nat Publ Gr* 2017, DOI: 10.1038/nrg.2017.63.
Hamady M, Lozupone C, Knight R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* 2010;**4**:17–27.
Hamm PS, Caimi NA, Northup DE *et al.* Western Bats as a Reservoir of Novel Streptomyces Species with Antifungal Activity. *Appl Environ Microbiol* 2017;**83**:3057.
Hanna JM. Climate, altitude, and blood pressure. *Hum Biol* 1999;**71**:553–82.
Hansen J, Gulati A, Sartor RB. The role of mucosal immunity and host genetics in defining intestinal commensal bacteria. *Curr Opin Gastroenterol* 2010;**26**:564–71.
Hardham JM, King KW, Dreier K *et al.* Transfer of Bacteroides splanchnicus to Odoribacter gen. nov. as Odoribacter splanchnicus comb. nov., and description of Odoribacter denticanis sp. nov., isolated from the crevicular spaces of canine periodontitis patients. *Int J Syst Evol Microbiol* 2008;**58**:103–9.
Harland PSEG. Skeletal variation in wild house mice from Peru. *Ann Mag Nat Hist* 1958;**1**:193–6.
Hayashi H, Takahashi R, Nishi T *et al.* Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *J Med Microbiol* 2005;**54**:1093–101.
He G, Shankar R a, Chzhan M *et al.* Noninvasive measurement of anatomic structure and intraluminal oxygenation in the gastrointestinal tract of living mice with spatial and spectral EPR imaging. *Proc Natl Acad Sci U S A* 1999;**96**:4586–91.
Heijtz RD, Wang S, Anuar F *et al.* Normal gut microbiota modulates brain development and behavior. *Proc Natl Acad Sci U S A* 2011;**108**:3047–52.
Herwig RP, Staley JT, Nerini MK *et al.* Baleen whales: preliminary evidence for forestomach microbial fermentation. *Appl Environ Microbiol* 1984;**47**:421–3.
Hijmans RJ, Cameron SE, Parra JL *et al.* Very high resolution interpolated climate surfaces for global land areas. *Int J Clim* 2005;**25**:1965–78.
Hooper L V, Midtvedt T, Gordon JI. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu Rev Nutr*

2002;**22**:283–307.
Hoyt JR, Cheng TL, Langwig KE *et al.* Bacteria isolated from bats inhibit the growth of Pseudogymnoascus destructans, the causative agent of white-nose syndrome. *PLoS One* 2015;**10**:1–12.
Hume ID. Marsupial Nutrition. *Cambridge, UK Cambridge Univ Press* 1999.
Hungate RE. The culture of Eudiplodinium neglectum with experiments on the digestion of cellulose. *Biol Bull* 1942;**83**:303–19.
Hungate RE. Further experiments on cellulose digestion by the protozoa in the rumen of cattle. *Biol Bull* 1943;**84**:157–63.
Hungate RE. Studies on cellulose fermentation. III. The culture and isolation for cellulose-decomposing bacteria from the rumen of cattle. *J Bacteriol* 1947;**53**:631–45.
Hungate RE. The rumen and its microbes. *Acad Press New York, NY* 1966.
Huttenhower C, Fah Sathirapongsasuti J, Segata N *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* 2012;**486**:207–14.
Irvin MR, Wineinger NE, Rice TK *et al.* Genome-wide detection of allele specific copy number variation associated with insulin resistance in african americans from the hyperGEN study. *PLoS One* 2011;**6**:9–11.
James AG, Austin CJ, Cox DS *et al.* Microbiological and biochemical origins of human axillary odour. *FEMS Microbiol Ecol* 2013;**83**:527–40.
Jochmans-Lemoine A, Shahare M, Soliz J *et al.* HIF1α and physiological responses to hypoxia are correlated in mice but not in rats. *J Exp Biol* 2016;**219**:3952–61.
Jochmans-Lemoine A, Villalpando G, Gonzales M *et al.* Divergent physiological responses in laboratory rats and mice raised at high altitude. *J Exp Biol* 2015;**218**:1035–43.
Jousset A, Bienhold C, Chatzinotas A *et al.* Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J* 2017;**11**:853–62.
Jumpertz R, Le DS, Turnbaugh PJ *et al.* Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *Am J Clin Nutr* 2011;**94**:58–65.
Kaushik S, Sowdhamini R. Distribution, classification, domain architectures and evolution of prolyl oligopeptidases in prokaryotic lineages. *BMC Genomics* 2014;**15**:1–13.
Kikuchi Y, Hayatsu M, Hosokawa T *et al.* Symbiont-mediated insecticide resistance. *Proc Natl Acad Sci* 2012;**109**:8618–22.
Knight R, Callewaert C, Marotz C *et al.* The Microbiome and Human

Biology. *Annu Rev Genomics Hum Genet* 2017;**183**:65–86.
Knights D, Silverberg MS, Weersma RK *et al.* Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med* 2014;**6**:107.
Kohl KD, Dearing DM. The woodrat gut microbiota as an experimental system for understanding microbial metabolism of dietary toxins. *Front Microbiol* 2016;**7**:1–9.
Kohl KD, Dearing MD, van Dam N. Experience matters: prior exposure to plant toxins enhances diversity of gut microbes in herbivores. *Ecol Lett* 2012;**15**:1008–15.
Kohl KD, Miller AW, Marvin JE *et al.* Herbivorous rodents (Neotoma spp.) harbour abundant and active foregut microbiota. *Environ Microbiol* 2014a;**16**:2869–78.
Kohl KD, Samuni-Blank M, Lymberakis P *et al.* Effects of Fruit Toxins on Intestinal and Microbial β-Glucosidase Activities of Seed-Predating and Seed-Dispersing Rodents (Acomys spp.). *Physiol Biochem Zool* 2016;**89**:198–205.
Kohl KD, Stengel A, Dearing MD. Inoculation of tannin-degrading bacteria into novel hosts increases performance on tannin-rich diets. *Environ Microbiol* 2016;**18**:1720–9.
Kohl KD, Weiss RB, Cox J *et al.* Gut microbes of mammalian herbivores facilitate intake of plant toxins. van Dam N (ed.). *Ecol Lett* 2014b;**17**:1238–46.
Kovatcheva-Datchary P, Nilsson A, Akrami R *et al.* Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. *Cell Metab* 2015;**22**:971–82.
Kraus DM, Elliott GS, Chute H *et al.* CSMD1 Is a Novel Multiple Domain Complement-Regulatory Protein Highly Expressed in the Central Nervous System and Epithelial Tissues. *J Immunol* 2006;**176**:4419–30.
Kreisinger J, Bastien G, Hauffe HC *et al.* Interactions between multiple helminths and the gut microbiota in wild rodents. *Philos Trans R Soc B Biol Sci* 2015;**370**:20140295.
Kreisinger J, Cížková D, Vohánka J *et al.* Gastrointestinal microbiota of wild and inbred individuals of two house mouse subspecies assessed using high throughput parallel pyrosequencing. *Mol Ecol* 2014, DOI: 10.1111/mec.12909.
Kuhn F, Natsch A. Body odour of monozygotic human twins: a common pattern of odorant carboxylic acids released by a bacterial aminoacylase from axilla secretions contributing to an inherited body odour type. *J R Soc Interface* 2009;**6**:377–92.

Lan D, Ji W, Lin B *et al.* Correlations between gut microbiota community structures of Tibetans and geography. *Sci Rep* 2017;**7**:1–9.
Langille MGI, Zaneveld J, Caporaso JG *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 2013;**31**:814–21.
Lay C, Rigottier-Gois L, Holmstrøm K *et al.* Colonic microbiota signatures across five northern European countries. *Appl Environ Microbiol* 2005;**71**:4153–5.
Leamy LJ, Kelly S a, Nietfeldt J *et al.* Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biol* 2014a;**15**:552.
Leamy LJ, Kelly SA, Nietfeldt J *et al.* Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biol* 2014b;**15**:552.
Leclaire S, Nielsen JF, Drea CM. Bacterial communities in meerkat anal scent secretions vary with host sex, age, and group membership. *Behav Ecol* 2014;**25**:996–1004.
Lender N, Talley NJ, Enck P *et al.* Review article: Associations between Helicobacter pylori and obesity - An ecological study. *Aliment Pharmacol Ther* 2014;**40**:24–31.
Ley RE, Bäckhed F, Turnbaugh P *et al.* Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* 2005;**102**:11070–5.
Ley RE, Hamady M, Lozupone C *et al.* Evolution of mammals and their gut microbes. *Science* 2008a;**320**:1647–51.
Ley RE, Lozupone CA, Hamady M *et al.* Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* 2008b;**6**:776–88.
Ley RE, Turnbaugh PJ, Klein S *et al.* Microbial ecology: human gut microbes associated with obesity. *Nature* 2006;**444**:1022–3.
Leyden JJ, McGinley KJ, Hölzle E *et al.* The Microbiology of the Human Axilla and Its Relationship to Axillary Odor. *J Invest Dermatol* 1981;**77**:413–6.
Li D, Chen B, Zhang L *et al.* The musk chemical composition and microbiota of Chinese forest musk deer males. *Sci Rep* 2016a;**6**:18975.
Li H, Li T, Beasley DAE *et al.* Diet diversity is associated with beta but not alpha diversity of pika gut microbiota. *Front Microbiol* 2016b;**7**:1–9.
Li H, Li T, Yao M *et al.* Pika gut may select for rare but diverse environmental bacteria. *Front Microbiol* 2016c;**7**:1–11.
Li K, Dan Z, Gesang L *et al.* Comparative analysis of gut microbiota of

native Tibetan and Han populations living at different altitudes. *PLoS One* 2016d;**11**:1–16.
Li L, Zhao X. Comparative analyses of fecal microbiota in Tibetan and Chinese Han living at low or high altitude by barcoded 454 pyrosequencing. *Sci Rep* 2015;**5**:1–10.
Li Q, Korzan WJ, Ferrero DM *et al.* Synchronous Evolution of an Odor Biosynthesis Pathway and Behavioral Response. *Curr Biol* 2012:11–20.
Linnenbrink M, Wang J, Hardouin E a *et al.* The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Mol Ecol* 2013;**22**:1904–16.
Liu CT, Monda KL, Taylor KC *et al.* Genome-Wide Association of Body Fat Distribution in African Ancestry Populations Suggests New Loci. *PLoS Genet* 2013;**9**, DOI: 10.1371/journal.pgen.1003681.
Liu Z, Geboes K, Heremans H *et al.* Role of interleukin-12 in the induction of mucosal inflammation and abrogation of regulatory T cell function in chronic experimental colitis. *Eur J Immunol* 2001;**31**:1550–60.
Lowy F. Stapylococcus aureus infections. *N Engl J Med* 1998;**339**:520–32.
Lu H-P, Lai Y-C, Huang S-W *et al.* Spatial heterogeneity of gut microbiota reveals multiple bacterial communities with distinct characteristics. *Sci Rep* 2014;**4**:6185.
Lundberg R, Toft MF, August B *et al.* Antibiotic-treated versus germ-free rodents for microbiota transplantation studies. *Gut Microbes* 2016;**7**:68–74.
Lynch C. Clinal variation in cold adaptation in Mus domesticus: verification of predictions from laboratory populations. *Am Nat* 1992;**139**:1219–36.
Ma J, Coarfa C, Qin X *et al.* mtDNA haplogroup and single nucleotide polymorphisms structure human microbiome communities. *BMC Genomics* 2014;**15**:257.
Ma X. TNF-α and IL-12:a balancing act in macrophage functioning. *Microbes Infect* 2001;**3**:121–9.
Margulis L. Symbiogenesis and symbionticism. In: Margulis L, Fester R, editors. Symbiosis As a Source of Evolutionary Innovation: Speciation and Morphogenesis. *MIT Press Cambridge, MA* 1991:1–14.
Mariat D, Firmesse O, Levenez F *et al.* The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol* 2009;**9**:123.
Martín R, Bermúdez-Humarán LG, Langella P. Gnotobiotic Rodents: An In Vivo Model for the Study of Microbe–Microbe Interactions. *Front Microbiol* 2016;**7**:1–7.
Masella AP, Bartram AK, Truszkowski JM *et al.* PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 2012;**13**:31.

Matamoros S, Gras-Leguen C, Le Vacon F *et al.* Development of intestinal microbiota in infants and its impact on health. *Trends Microbiol* 2013;**21**:167–73.
Maurice CF, Cl Knowles S, Ladau J *et al.* Marked seasonal variation in the wild mouse gut microbiota. *ISME J* 2015;**9**:1–12.
McCann JC, Wickersham TA, Loor JJ. High-throughput Methods Redefine the Rumen Microbiome and Its Relationship with Nutrition and Metabolism. 2014;**8**:109–25.
McFall-Ngai M, Hadfield MG, Bosch TCG *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci* 2013;**110**:3229–36.
McKnite AM, Perez-Munoz ME, Lu L *et al.* Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PLoS One* 2012;**7**:e39191.
Merritt GC, Goodrich BS, Hesterman ER *et al.* Microflora and volatile fatty acids present in inguinal pouches of the wild rabbit,Oryctolagus cuniculus, in Australia. *J Chem Ecol* 1982;**8**:1217–25.
Meyer F, Paarmann D, D'Souza M *et al.* The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 2008;**9**:386.
Miller AW, Kohl KD, Dearing DM. The gastrointestinal tract of the white-throated woodrat (Neotoma albigula) harbors distinct consortia of oxalate-degrading bacteria. *Appl Environ Microbiol* 2014;**80**:1595–601.
Moeller AHA, Foerster S, Wilson ML *et al.* Social behavior shapes the chimpanzee pan-microbiome. *Sci Adv* 2016a;**2**:e1500997–e1500997.
Moeller AH, Caro-Quintero A, Mjungu D *et al.* Cospeciation of gut microbiota with hominids. *Science (80-)* 2016b;**353**:380–2.
Moeller AH, Degnan PH, Pusey AE *et al.* Chimpanzees and humans harbour compositionally similar gut enterotypes. *Nat Commun* 2012;**3**:1179.
Moeller AH, Peeters M, Ayouba A *et al.* Stability of the gorilla microbiome despite simian immunodeficiency virus infection. *Mol Ecol* 2015;**24**:690–7.
Moeller AH, Peeters M, Ndjango JB *et al.* Sympatric chimpanzees and gorillas harbor convergent gut microbial communities. *Genome Res* 2013a;**23**:1715–20.
Moeller AH, Shilts M, Li Y *et al.* Siv-induced instability of the chimpanzee gut microbiome. *Cell Host Microbe* 2013b;**14**:340–5.
Moeller AH, Suzuki TA, Lin D *et al.* Dispersal limitation promotes the diversification of the mammalian gut microbiota. *Proc Natl Acad Sci* 2017;**114**:201700122.

Moran NA, Sloan DB. The Hologenome Concept: Helpful or Hollow? *PLoS Biol* 2015;**13**:1–10.
Moreau NM, Goupry SM, Antignac JP *et al.* Simultaneous measurement of plasma concentrations and 13C-enrichment of short-chain fatty acids, lactic acid and ketone bodies by gas chromatography coupled to mass spectrometry. *J Chromatogr B Anal Technol Biomed Life Sci* 2003;**784**:395–403.
Moreno-Indias I, Torres M, Montserrat JM *et al.* Intermittent hypoxia alters gut microbiota diversity in a mouse model of sleep apnoea. *Eur Respir J* 2015;**45**:1055–65.
Mortensen F V, Nielsen H, Mulvany MJ *et al.* Short chain fatty acids dilate isolated human colonic resistance arteries. *Gut* 1990;**31**:1391–4.
Mowat AM, Agace WW. Regional specialization within the intestinal immune system. *Nat Rev Immunol* 2014;**14**:667–85.
Muegge BD, Kuczynski J, Knights D *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 2011;**332**:970–4.
Mueller S, Saunier K, Hanisch C *et al.* Differences in Fecal Microbiota in Different European Study Populations in Relation to Age , Gender , and Country : a Cross-Sectional Study Differences in Fecal Microbiota in Different European Study Populations in Relation to Age , Gender , and Country : *Appl Environ Microbiol* 2006, DOI: 10.1128/AEM.72.2.1027–1033.2006.
Murphy EF, Cotter PD, Healy S *et al.* Composition and energy harvesting capacity of the gut microbiota: Relationship to diet, obesity and time in mouse models. *Gut* 2010;**59**:1635–42.
Nelson TM, Rogers TL, Brown M V. The gut bacterial community of mammals from marine and terrestrial habitats. *PLoS One* 2013;**8**:1–8.
Nelson TM, Rogers TL, Carlini AR *et al.* Diet and phylogeny shape the gut microbiota of Antarctic seals: A comparison of wild and captive animals. *Environ Microbiol* 2013;**15**:1132–45.
Nemoto K, Osawa R, Hirota K *et al.* An investigation of gram-negative tannin-protein complex degrading bacteria in fecal flora of various mammals. *J Vet Med Sci* 1995;**57**:921–6.
Nguyen TLA, Vieira-Silva S, Liston A *et al.* How informative is the mouse for human gut microbiota research? *Dis Model Mech* 2015;**8**:1–16.
Nishida AH, Ochman H. Rates of Gut Microbiome Divergence in Mammals. *Mol Ecol* 2017;**12**:3218–21.
Nuriel-Ohayon M, Neuman H, Koren O. Microbial changes during pregnancy, birth, and infancy. *Front Microbiol* 2016;**7**:1–13.

Nutting CW, Islam S, Daugirdas JT. Vasorelaxant effects of short chain fatty acid salts in rat caudal artery. *Am J Physiol* 1991;**261**:H561-7.
Ochman H, Worobey M, Kuo C-H *et al.* Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biol* 2010;**8**:e1000546.
Oliver KM, Russell JA, Moran NA *et al.* Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc Natl Acad Sci U S A* 2003;**100**:1803–7.
Olszak T, An D, Zeissig S *et al.* Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T Cell Function. *Science (80-)* 2012:6–10.
Org E, Parks BW, Joo JWJ *et al.* Genetic and environmental control of host-gut microbiota interactions. *Genome Res* 2015;**25**:1558–69.
Osawa R. Formation of a clear zone on tannin-treated brain heart infusion agar by a Streptococcus sp. isolated from feces of koalas. *Appl Environ Microbiol* 1990;**56**:829–31.
Osawa R, Sly LI. Occurrence of tannin-protein complex degrading Streptococcus sp. in feces of various animals. *Syst Appl Microbiol* 1992;**15**:144–7.
Ott SJ, Musfeldt M, Timmis KN *et al.* In vitro alterations of intestinal bacterial microbiota in fecal samples during storage. *Diagn Microbiol Infect Dis* 2004;**50**:237–45.
Ou J, Carbonero F, Zoetendal EG *et al.* Diet, microbiota, and microbial metabolitesin colon cancer risk in rural Africans and African-Americans. *Am J Clin Nutr* 2013;**98**:111–20.
Ouwehand AC, Salminen S, Arvola T *et al.* Microbiota composition of the intestinal mucosa: association with fecal microbiota? *Microbiol Immunol* 2004;**48**:497–500.
Ouwerkerk JP, van der Ark KCH, Davids M *et al.* Adaptation of Akkermansia muciniphila to the oxic-anoxic interface of the mucus layer. *Appl Environ Microbiol* 2016;**82**:6983–93.
Ovsyannikova IG, Kennedy RB, O'Byrne M *et al.* Genome-wide association study of antibody response to smallpox vaccine. *Vaccine* 2012;**30**:4182–9.
Palmer C, Bik EM, DiGiulio DB *et al.* Development of the human infant intestinal microbiota. *PLoS Biol* 2007;**5**:1556–73.
Parronchi P, Romagnani P, Annunziato F *et al.* Type 1 T-helper cell predominance and interleukin-12 expression in the gut of patients with Crohn's disease. *Am J Pathol* 1997;**150**:823–32.
Phifer-Rixey M, Nachman MW. Insights into mammalian biology from the

wild house mouse Mus musculus. *Elife* 2015;**2015**:1–13.
Phillips CD, Phelan G, Dowd SE *et al.* Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. *Mol Ecol* 2012;**21**:2617–27.
Piaggi P, Masindova I, Muller YL *et al.* A Genome-Wide Association Study Using a Custom Genotyping Array Identifies Variants in GPR158 Associated With Reduced Energy Expenditure in American Indians. *Diabetes* 2017;**66**:2284–95.
Piertney SB, Oliver MK. The evolutionary ecology of the major histocompatibility complex. *Heredity (Edinb)* 2006;**96**:7–21.
Pluznick J. A novel SCFA receptor, the microbiota, and blood pressure regulation. *Gut Microbes* 2014;**5**:202–7.
Pluznick JL, Protzko RJ, Gevorgyan H *et al.* Olfactory receptor responding to gut microbiota- derived signals plays a role in renin secretion and blood pressure regulation. *Proc Natl Acad Sci* 2013;**110**:4410–5.
Pluznick JL, Zou DJ, Zhang X *et al.* Functional expression of the olfactory signaling system in the kidney. *Proc Natl Acad Sci U S A* 2009;**106**:2059–64.
Ppatil D, Pdhotre D, Gchavan S *et al.* Molecular analysis of gut microbiota in obesity among Indian individuals. *J Biosci* 2012;**37**:647–57.
Preston FW. The Canonical Distribution of Commonness and Rarity: Part I. *Ecology* 1962;**43**:185.
Price MN, Dehal PS, Arkin AP. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009;**26**:1641–50.
Qiu Q, Zhang G, Ma T *et al.* The yak genome and adaptation to life at high altitude. *Nat Genet* 2012;**44**:946–9.
Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**:590–6.
Rahat-Rozenbloom S, Fernandes J, Gloor GB *et al.* Evidence for greater production of colonic short-chain fatty acids in overweight than lean humans. *Int J Obes* 2014;**38**:1525–31.
Rawls JF, Mahowald M a, Ley RE *et al.* Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* 2006;**127**:423–33.
Reunanen J, Kainulainen V, Huuskonen L *et al.* Akkermansia muciniphila adheres to enterocytes and strengthens the integrity of the epithelial cell layer. *Appl Environ Microbiol* 2015;**81**:3655–62.
Richardson AJ. Simultaneous determination of volatile and non-volatile

acidic ferment. *Lett Appl Microbiol* 1989;**9**:5–8.
Rideout JR, He Y, Navas-Molina J a *et al.* Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* 2014;**2**:e545.
Rivera-Chávez F, Zhang LF, Faber F *et al.* Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella. *Cell Host Microbe* 2016;**19**:443–54.
Rivière G, Michaud A, Corradi HR *et al.* Characterization of the first angiotensin-converting like enzyme in bacteria: Ancestor ACE is already active. *Gene* 2007;**399**:81–90.
Roberts DF. Body weight, race and climate. *Am J Phys Anthropol* 1953;**11**:533–58.
Robertson BR, O'Rourke JL, Neilan BA *et al.* Mucispirillum schaedleri gen. nov., sp. nov., a spiral-shaped bacterium colonizing the mucus layer of the gastrointestinal tract of laboratory rodents. *Int J Syst Evol Microbiol* 2005;**55**:1199–204.
Rollins-Smith LA, Ramsey JP, Pask JD *et al.* Amphibian immune defenses against chytridiomycosis: Impacts of changing environments. *Integr Comp Biol* 2011;**51**:552–62.
Rosenberg E, Koren O, Reshef L *et al.* The role of microorganisms in coral health, disease and evolution. *Nat Rev Microbiol* 2007;**5**:355–62.
Rosshart SP, Vassallo BG, Angeletti D *et al.* Wild Mouse Gut Microbiota Promotes Host Fitness and Improves Disease Resistance. *Cell* 2017:1–14.
Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 2009a;**9**:313–23.
Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 2009b;**9**:313–23.
Sahu NP, Kamra DN. Microbial eco-system of the gastro-intestinal tract of wild herbivorous animals. *J Appl Anim Res* 2002;**21**:207–30.
Salas JT, Chang TL. Microbiome in human immunodeficiency virus infection. *Clin Lab Med* 2014;**34**:733–45.
Salgado-Flores A, Hagen LH, Ishaq SL *et al.* Rumen and cecum microbiomes in reindeer (Rangifer tarandus tarandus) are changed in response to a lichen diet and may affect enteric methane emissions. *PLoS One* 2016;**11**:1–22.
Sampson TR, Mazmanian SK. Control of brain development, function, and behavior by the microbiome. *Cell Host Microbe* 2015;**17**:565–76.
Samuel BS, Shaito A, Motoike T *et al.* Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-

coupled receptor, Gpr41. *Proc Natl Acad Sci* 2008;**105**:16767–72.
Sanders JG, Beichman AC, Roman J *et al.* Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. *Nat Commun* 2015;**6**:8285.
Sasaki E, Shimada T, Osawa R *et al.* Isolation of tannin-degrading bacteria isolated from feces of the Japanese large wood mouse, Apodemus speciosus, feeding on tannin-rich acorns. *Syst Appl Microbiol* 2005;**28**:358–65.
Schirmer M, Smeekens SP, Vlamakis H *et al.* Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* 2016;**167**:1125–36.
Schnorr SL, Candela M, Rampelli S *et al.* Gut microbiome of the Hadza hunter-gatherers. *Nat Commun* 2014;**5**:3654.
Schwiertz A, Taras D, Schäfer K *et al.* Microbiota and SCFA in lean and overweight healthy subjects. *Obesity (Silver Spring)* 2010;**18**:190–5.
Shah HN, Collins DM. Prevotella, a new genus to include Bacteroides melaninogenicus and related species formerly classified in the genus Bacteroides. *Int J Syst Bacteriol* 1990;**40**:205–8.
Shimada T, Saitoh T, Sasaki E *et al.* Role of tannin-binding salivary proteins and tannase-producing bacteria in the acclimation of the Japanese wood mouse to acorn tannins. *J Chem Ecol* 2006;**32**:1165–80.
Sijpesteijn KA. Cellulose-decomposing bacteria from the rumen of cattle. *PhD Thesis, Leiden Univ Eduard Ijdo NV, Leiden* 1948.
Simůnek J, Hodrová B, Bartonová H *et al.* Chitinolytic bacteria of the mammal digestive tract. *Folia Microbiol (Praha)* 2001;**46**:76–8.
Sin YW, Buesching CD, Burke T *et al.* Molecular characterization of the microbial communities in the subcaudal gland secretion of the European badger (Meles meles). *FEMS Microbiol Ecol* 2012;**81**:648–59.
Smith B, Li N, Andersen AS *et al.* Optimising bacterial DNA extraction from faecal samples: comparison of three methods. *Open Microbiol J* 2011;**5**:14–7.
Van Soest PJ. Nutritional ecology of the ruminant. *Cornell Univ Press Ithaca, NY* 1994.
Sonoyama K, Fujiwara R, Takemura N *et al.* Response of gut microbiota to fasting and hibernation in Syrian hamsters. *Appl Environ Microbiol* 2009;**75**:6451–6.
Sparks MA, Crowley SD, Gurley SB *et al.* Classical Renin-Angiotensin system in kidney physiology. *Compr Physiol* 2014;**4**:1201–28.
Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* 2011;**9**:279–

90.
Srinivas G, Möller S, Wang J *et al.* Genome-wide mapping of gene–microbiota interactions in susceptibility to autoimmune skin blistering. *Nat Commun* 2013;**4**:2462.
Stearns JC, Lynch MDJ, Senadheera DB *et al.* Bacterial biogeography of the human digestive tract. *Sci Rep* 2011;**1**:170.
Stevens, C.E. HID. Comparative physiology of the vertebrate digestive system. 1995.
Stevens CE, Hume ID. Comparative Physiology of the Vertebrate Digestive System. *Cambridge, UK Cambridge Univ Press* 1995.
Stevenson TJ, Duddleston KN, Buck CL. Effects of season and host physiological state on the diversity, density, and activity of the arctic ground squirrel cecal microbiota. *Appl Environ Microbiol* 2014;**80**:5611–22.
Storz JF. Hemoglobin Function and Physiological Adaptation to Hypoxia in High-Altitude Mammals. *J Mammal* 2007;**88**:24–31.
Storz JF, Baze M, Waite JL *et al.* Complex signatures of selection and gene conversion in the duplicated globin genes of house mice. *Genetics* 2007;**177**:481–500.
Strobel HJ. Vitamin B12-dependent propionate production by the ruminal bacterium Prevotella ruminicola 23. *Appl Environ Microbiol* 1992;**58**:2331–3.
Sugihara Y, Kawasaki A, Tsujimoto Y *et al.* Potencies of phosphine peptide inhibitors of mammalian thimet oligopeptidase and neurolysin on two bacterial pz peptidases. *Biosci Biotechnol Biochem* 2007;**71**:594–7.
Sun B, Wang X, Bernstein S *et al.* Marked variation between winter and spring gut microbiota in free-ranging Tibetan Macaques (Macaca thibetana). *Sci Rep* 2016;**6**:26035.
Suzuki TA. Links between Natural Variation in the Microbiome and Host Fitness in Wild Mammals. *Integr Comp Biol* 2017;**57**:756–69.
Suzuki TA, Nachman MW. Spatial heterogeneity of gut microbial composition along the gastrointestinal tract in natural populations of house mice. *PLoS One* 2016;**11**:1–15.
Suzuki T, Worobey M. Geographical variation of human gut microbial composition. *Biol Lett* 2014;**10**:20131037.
Tappeiner H. Untersuchungen über die gärung der cellulose insbesondere über deren lösung im darmkanale. *Z Biol* 1884;**20**:52–134.
Theis KR, Schmidt TM, Holekamp KE. Evidence for a bacterial mechanism for group-specific social odors among hyenas. *Sci Rep* 2012;**2**:1–8.
Theis KR, Venkataraman A, Dycus J a *et al.* Symbiotic bacteria appear to

mediate hyena social odors. *Proc Natl Acad Sci U S A* 2013;**110**:19832–7.
Thompson LR, Sanders JG, McDonald D *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 2017;**551**:457–63.
Tilman D, Downing JA. Biodiversity and stability in grasslands. *Nature* 1994;**367**:363–5.
Topping DL, Clifton PM. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiol Rev* 2001;**81**:1031–64.
Trinchieri G. Interleukin-12: a cytokine at the interface of inflammation and immunity. *Adv Immunol* 1998;**70**:83–243.
Tung J, Barreiro LB, Burns MB *et al.* Social networks predict gut microbiome composition in wild baboons. *Elife* 2015;**2015**:e05224.
Turnbaugh PJ, Hamady M, Yatsunenko T *et al.* A core gut microbiome in obese and lean twins. *Nature* 2009;**457**:480–4.
Turnbaugh PJ, Ley RE, Mahowald M a *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006;**444**:1027–31.
Turpin W, Espin-garcia O, Xu W *et al.* Association of host genome with intestinal microbial composition in a large healthy cohort. 2016;**48**, DOI: 10.1038/ng.3693.
Upala S, Jaruvongvanich V, Riangwiwat T *et al.* Association between Helicobacter pylori infection and metabolic syndrome: a systematic review and meta-analysis. *J Dig Dis* 2016;**17**:433–40.
Vieira FG, Lassalle F, Korneliussen TS *et al.* Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biol J Linn Soc* 2016;**117**:139–49.
Voigt CC, Caspers B, Speck S. Bats, Bacteria, and Bat Smell: Sex-Specific Diversity of Microbes in a Sexually Selected Scent Organ. *J Mammal* 2005;**86**:745–9.
Vujkovic-Cvijin I, Dunham RM, Iwai S *et al.* Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism. *Sci Transl Med* 2013;**5**:193ra91.
Wang J, Kalyan S, Steck N *et al.* Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nat Commun* 2015;**6**:6440.
Wang J, Linnenbrink M, Künzel S *et al.* Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild mice. *Proc Natl Acad Sci U S A* 2014;**111**:E2703-10.

Wang J, Thingholm LB, Skiecevic J *et al.* Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. 2016;**48**, DOI: 10.1038/ng.3695.
Wang T, Cai G, Qiu Y *et al.* Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J* 2012;**6**:320–9.
Warnecke L, Turner JM, Bollinger TK *et al.* Inoculation of bats with European Geomyces destructans supports the novel pathogen hypothesis for the origin of white-nose syndrome. *Proc Natl Acad Sci* 2012;**109**:6999–7003.
Wasielewski H, Alcock J, Aktipis A. Resource conflict and cooperation between human host and gut microbiota: implications for nutrition and health. *Ann N Y Acad Sci* 2016;**1372**:20–8.
Wedekind C, Füri S. Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proc Biol Sci* 1997;**264**:1471–9.
Wedekind C, Seebeck T, Bettens F *et al.* MHC-Dependent Mate Preferences in Humans. *Proc R Soc B Biol Sci* 1995;**260**:245–9.
Weese JS, Nichols J, Jalali M *et al.* The rectal microbiota of cats infected with feline immunodeficiency virus infection and uninfected controls. *Vet Microbiol* 2015;**180**:96–102.
Wei C, Wang H, Liu G *et al.* Genome-wide analysis reveals adaptation to high altitudes in Tibetan sheep. *Sci Rep* 2016;**6**:1–11.
Weisfeld GE, Czilli T, Phillips KA *et al.* Possible olfaction-based mechanisms in human kin recognition and inbreeding avoidance. *J Exp Child Psychol* 2003;**85**:279–95.
Weldon L, Abolins S, Lenzi L *et al.* The gut microbiota of wild mice. *PLoS One* 2015;**10**:1–15.
Whitaker JO, Dannelly HK, Prentice DA. Chitinase in Insectivorous Bats. *Source J Mammal* 2004;**85**:15–8.
Whitman WB, Rainey F, Kämpfer P *et al.* eds. *Bergey's Manual of Systematics of Archaea and Bacteria*. Chichester, UK: John Wiley & Sons, Ltd, 2015.
Williams B, Landay A, Presti RM. Microbiome alterations in HIV infection a review. *Cell Microbiol* 2016;**18**:645–51.
Wu GD, Chen J, Hoffmann C *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* 2011;**334**:105–8.
Yang T, Zubcevic J. Gut-brain axis in regulation of blood pressure. *Front Physiol* 2017;**8**:1–12.
Yasuda K, Oh K, Ren B *et al.* Biogeography of the Intestinal Mucosal and

Lumenal Microbiome in the Rhesus Macaque. *Cell Host Microbe* 2015;**17**:385–91.
Yatsunenko T, Rey FE, Manary MJ *et al.* Human gut microbiome viewed across age and geography. *Nature* 2012;**486**:222–7.
Zechman JM, Martin IG, Wellington JL *et al.* Perineal scent gland of wild and domestic cavies: Bacterial activity and urine as sources of biologically significant odors. *Physiol Behav* 1984;**32**:269–74.
Zeitouni NE, Chotikatum S, von Köckritz-Blickwede M *et al.* The impact of hypoxia on intestinal epithelial cell functions: consequences for invasion by bacterial pathogens. *Mol Cell Pediatr* 2016;**3**:14.
Zhang C, Zhang M, Wang S *et al.* Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. *ISME J* 2010;**4**:232–41.
Zhang Z, Geng J, Tang X *et al.* Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J* 2014;**8**:881–93.
Zhang Z, Xu D, Wang L *et al.* Convergent Evolution of Rumen Microbiomes in High-Altitude Mammals. *Curr Biol* 2016;**26**:1873–9.
Zhao J, Yao Y, Li D *et al.* Characterization of the Gut Microbiota in Six Geographical Populations of Chinese Rhesus Macaques (Macaca mulatta), Implying an Adaptation to High-Altitude Environment. *Microb Ecol* 2018, DOI: 10.1007/s00248-018-1146-8.
Zheng L, Kelly CJ, Colgan SP. Physiologic hypoxia and oxygen homeostasis in the healthy intestine. A Review in the Theme: Cellular Responses to Hypoxia. *Am J Physiol - Cell Physiol* 2015;**309**:C350–60.
Zhu L, Wu Q, Dai J *et al.* Evidence of cellulose metabolism by the giant panda gut microbiome. *Proc Natl Acad Sci U S A* 2011;**108**:17714–9.
Ziętak M, Kovatcheva-Datchary P, Markiewicz LH *et al.* Altered Microbiota Contributes to Reduced Diet-Induced Obesity upon Cold Exposure. *Cell Metab* 2016;**23**:1216–23.
Zilber-Rosenberg I, Rosenberg E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol Rev* 2008;**32**:723–35.
Zuntz N. Gesichtspunkte zum kritischen Studium der neueren Arbeiten auf dem Gebiete der Ernahrung. *Landwirtsch Jahrbücher* 1879;**8**:65–117.