Ecological forces shaping gut microbiota in wild animal populations

Tiantian Ren
Shangqiu, Henan, China

B.S., Wuhan University, 2010

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Biology

University of Virginia
August, 2016

Abstract

Trillions of bacteria colonize the digestive tracts of animals and form complex communities known as gut microbiota. Gut microbiota play important roles in diverse aspects of host biology, including nutrition, immune system development and behavior. Changes in gut microbial composition have been linked to a plethora of health and disease states. Previous studies on human and laboratory animals have shown that host diet, age, sex, genetics, environmental exposure all drive normal gut microbial variation. However, the ecological forces that shape gut microbial community structure in wild animal populations remain largely unknown. Studying gut microbiota in host's natural environment is crucial because it is where the actual actions of microbes-host coevolution take place. In this dissertation, I investigated the ecological forces shaping gut microbial communities in three wild animal populations, with each tackling the problem from a different angle. First, I conducted a longitudinal and cross-sectional study of gut microbiota in a well-studied population of wild baboons. Baboon gut microbiota were typical of omnivorous primates, and host age and diet had strong effects on gut microbial composition. Strikingly, baboon gut microbiota appeared to be highly dynamic such that samples collected from the same individual only a few days apart were as different from each other as samples collected over 10 years apart. Next, I conducted a comprehensive study of the temporal and spatial dynamics of gut microbiota of a well-studied red squirrel population. Red squirrels represent a very attractive system for studying gut microbial biogeography because they are territorial and experience strong seasonal fluctuations in their environment. This study revealed significant spatial patterns and

seasonal rhythm of gut microbiota within a host population. Beyond the environmental effects, I also found evidences for individuality and maternal effect in red squirrel gut microbiota. Lastly, to test whether and how host adaptive radiations structure the composition of microbial communities, I investigated the gut microbiota of seven species of *Anolis* lizards belonging to three "ecomorphs" on Puerto Rico and Florida. Our results indicate that gut microbial communities are only weakly shaped by the diversification of their lizard hosts due to the strikingly high levels of microbial variation observed within *Anolis* species. In summary, I have characterized the temporal, spatial and phylogenetic patterns of gut microbiota in three different wild animal hosts, and identified various environmental and host factors underlying these patterns. These findings may contribute towards a better understanding of how ecological processes govern the gut microbial diversity in natural environment.

Acknowledgement

"Horse month of the Monkey year", or Hou Nian Ma Yue, is a commonly used Chinese slang that refers to a date far in the unforeseeable future, or doing something with little chance of success. When I first arrived in US and began my Ph.D. journey 6 years ago, I thought I would not finish it until Hou Nian Ma Yue. Coincidentally and happily, the time is coming for real in June (Horse month) 2016 (Monkey year)! This long journey is full of ups and downs, challenges and rewards, tears and excitement, and is an unforgettable and valuable experience that have made me grow stronger. I owe my gratitude to all those people who have made this dissertation possible and made this journey an experience I will cherish forever.

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Martin Wu for your continuous support and encouragement during my Ph.D. study and research. Thank you for your guidance and great effort that you have put into training me in the scientific field, from troubleshooting to paper writing. Thank you for helping me with the late night experiments. Thank you for being patient with me, and understanding my “crazy” ideas. Your creative ideas and critical thinking during our discussions always inspire me and bring me to a higher level of thinking. What I have learned from you is far beyond research because you have taught me how to view this world from a new perspective. I feel lucky to study under the same cherry blossom trees as you did 20 years ago, and I feel lucky to work with you when I travel overseas to pursue my Ph.D. at UVA.

I also would like to thank the other members of my committee for the assistance they provided at all levels of my research project. Special thanks to Dr. Edmund D. Brodie III, for taking time out from your busy schedule to serve as my first reader. Your expertise and insightful comments were invaluable for me in completing this work. I want to thank Dr. Janis Antonovics for the inspiring questions during our meetings, which have always made me think deeper and broader. I will forever remember Darwin's birthday. I would like to acknowledge Dr. Robert Cox, for taking time to answer my questions whenever I asked, and for the enjoyable collaboration on Chapter 4. Thank you, Bob, for showing me the art of science through the figures you made, which were both beautiful and scientifically sound. Finally, I would like to thank Dr. Jason Papin from the Department of Biomedical Engineering for serving as my external reader, and being supportive and understanding through my whole Ph.D. research. It's my honor to have such a great committee.

My sincere thanks also go to collaborators of my projects. Dr. Elizabeth A. Archie and Laura E. Grieneisen contributed to sample collection, data analysis and manuscript preparation of baboon project (Chapter 2). Drs. Andrew McAdam and Ben Dantzer provided valuable samples, data and critical feedbacks for the red squirrel project (Chapter 3). Dr. Robert Cox and Ariel F. Kahrl contributed to sample collection, data analysis and manuscript preparation of lizard project (Chapter 4). Drs. Michele Sale, Birgit Winther, Stephen V. Early, Emma Kaitlynn Allen and Tam Nhu Nguyen collected samples and provided helpful comments on the manuscript for the adenoid project (Appendix 1).

I am deeply grateful to Dr. Anthony Spano, who offered generous help and support throughout my study. Thank you for teaching me the techniques of chromatography, English and life. And many thanks for always finding me everything I need (Your lab is like Doraemon's pocket!)

My gratitude also goes to AnhThu Nguyen, who went through the tedious program setting and troubleshooting with me, and always be there whenever I was struggling. I enjoyed our lunches and talks, and it was great to meet someone with whom I have endless conversation topics.

I am indebted to Dr. Alexander F. Koeppel, who has one of the kindest hearts and minds I have ever met. Thank you for taking time to teach me programming skills, and debugging with me. Thank you for sharing helpful career advice and writing a great recommendation letter for my job search.

To my past and current lab members, Drs. Zhang Wang and Xin Zhao, and future Drs. Cory Weller, Yingnan Gao and Yiyi Pu. Thank you for your help and support, and for making the lab work enjoyable. I wish you all the best in your future endeavors. In particular, I am grateful to Dr. Zhang Wang. As my academic "elder brother", you have given me the most helpful assistances in science and outside the academic life. You are the first one to help me settle down in this new environment and culture. To the undergraduates who worked together with me, Maurice Wong, Sarah Smith, Linda He

and Dominique Ulrike Glatt, thank you for your help with this work, and thank you for teaching me how to be a good mentor. Special thanks to Maurice, for being my English teacher (Do you still remember how many feet are in a mile?).

I would also like to express my gratitude to other departmental members. The guidance and encouragement from my rotation advisor Dr. Herman Wijnen and academic advisor Dr. Lei Li helped me survive the first year in graduate school. Dr. Christopher Deppmann and Dr. Jay Hirsh generously provided ideas and resources for my experiments. I also wish to thank my three great DGSs, Dr. Keith G. Kozminski, Dr. Dorothy Schafer, and Dr. Barry Condron for always being available. Working as a TA in Dr. David Kittlesen's class five times is a memorable experience, and thank you for teaching me how to be a good teacher. Ms. Kay Christopher, thank you for your help and patience in experiment preparation. Ms. Joanne Chaplin, you are my role model in everyway. It has been a pleasure to work with you. I also would like to thank the Department of Biology for the financial support.

Thanks to my fellow graduate students and EEbio community, for the support and feedback on experiments, manuscripts and talks. A very special thank goes out to my fellow Chinese students and colleagues, Dr. Yan Hu, Dr. Hai Liu, Dr. Ying Wang, Dr. Xiaozeng Yang, Chun Su, Dr. Dong Suo, Dr. Ashley Dai, Yanping Li and Dr. Zhijin Zhang. You made me feel like family while far away from home.

At last but by no means least, I would like to express my sincerest gratitude to my family. My parents are my greatest inspiration and source of encouragement through my entire life. Highest regards to my father, Yiming Ren, who has been my hero all the time and taught me the value of hard working, persistence and kindness. Deepest gratitude to my mother, Haiyan Yu, who gave me the greatest gift in the world. You always believed in me. I would especially like to thank my beloved husband, Dr. Pei Wang, for being my best friend, my soul mate, and my home. Thank you for expanding my horizons, curing my “procrastination”, and making me a better me. I can never thank you enough. This work is dedicated to you.

Table of Contents

Chapter 1. Introduction

Gut Microbiota: A brief History

Ever since Robert Koch proved "germ theory", the view of host-associated microbes had been dominated by its role in causing infectious disease. However, a new view that most interactions between the host and microbes are not pathogenic but mutually beneficial emerged at the end of 19th century. As Theodore Escherich observed that "the apparently randomly appearing bacteria in normal feces and the intestinal tract" could interact and influence physiological properties of their host (Escherich, 1885), research on the gut microbiota (formerly 'the normal flora') was initiated. Before effective methods were developed for culturing anaerobic bacteria, researches were mainly focused on aerobic bacteria, which led many scientists to regard the *Escherichia coli*, a facultative bacterium named after Escherich, to be predominant in human gut (Savage, 2001). Development of anaerobic culture technology facilitated some important understandings of gut microbiota in mid 20th century, such as the discovery of dominant strict anaerobic species and early succession of gut microbial species in newborns (Schaedler *et al.*, 1965; Drasar, 1967). One major drawback of these early studies is that the vast majority of microbes are uncultivable, and therefore our understanding of the composition of gut microbiota is rudimentary and biased.

In the past decade, advancement of DNA sequencing technology has revolutionized the field of microbial ecology. The culture independent sequencing of the bacterial 16S rRNA (or small subunit ribosomal RNA) gene as a phylogenetic marker enabled characterization of microbial communities on an unprecedented scale. Indeed, in-depth survey of thousands of samples directly from environment has becomes a routine due to the continuous decrease in the cost of next generation sequencing technologies. As a result, study of gut microbiota has shifted from studying bacteria in isolation to studying bacteria in the context of the entire community. Researchers have discovered extraordinary level of gut microbial diversity and important roles

gut microbiota play in host health and disease (Cho and Blaser, 2012). Although many studies of gut microbiota are descriptive and exploratory, there are great opportunities and a growing interest to apply and test ecological principles in the field to understand the underlying processes that generate and maintain gut microbial diversity, and how emergent properties of the community influence host and respond to changes in environment (Christian *et al.*, 2015).

Gut microbiota: who are there and what do they do?

Despite the tremendous microbial diversity on our planet, animal intestinal tracts are dominated by only a few bacterial phyla (Firmicutes, Bacteroidetes, Actinobacteria and Proteobacteria). However, there is substantial diversity at lower taxonomic level (e.g. species) within these major divisions (Ley *et al.*, 2006). For example, it was estimated that each human harbors more than 1,000 "species-like" phylotypes (Claesson *et al.*, 2009). This fan-like phylogenetic architecture or microdiversity could result from recurring selective sweeps followed by genetic drifts (Bäckhed *et al.*, 2005; Ley *et al.*, 2006; Koeppel *et al.*, 2008; Bik *et al.*, 2010). Although the membership of main bacterial groups is relatively stable, their relative abundances vary remarkably both between individuals and within the same individual over time (Lozupone *et al.*, 2012).

Considered as an indispensable "organ", gut microbial community provides important ecosystem services to the host. Gut microbiota assist in the fermentation of indigestible dietary substrates and the synthesis of essential nutrients such as Vitamin A and D (Ley *et al.*, 2008; Qin *et al.*, 2010). Beyond their nutritional role, gut microbiota protect host by excluding pathogens (Hooper *et al.*, 2012) and contribute to the development of host immune system (Sommer and Bäckhed, 2013).

As reminiscent of the famous line "All happy families are alike; each unhappy family is unhappy in its own way", a growing body of evidence shows that many diseases are associated with imbalance of the gut microbial composition (known as "dysbiosis") and function. For instance, disrupted human microbiota have been linked to obesity (Ley *et al.*, 2005; Turnbaugh *et al.*, 2009), diabetes (Qin *et al.*, 2010; Vrieze *et al.*, 2012; Graessler *et al.*, 2013), cardiovascular disease (Wang *et al.*, 2011), inflammatory bowel disease (Frank *et al.*, 2007; Morgan *et al.*, 2012) and depression (Naseribafrouei *et al.*, 2014). Medical community has now recognized the potential to use gut microbiota as diagnostic and therapeutic targets (Hollister *et al.*, 2014).

Processes structuring gut microbial community

Understanding ecological processes that establish and maintain gut microbial diversity is an essential first step to understand how the "host-microbe ecosystem" functions. Two types of processes, deterministic and stochastic, influence the assembly of gut microbial community. Deterministic processes include environmental selection (abiotic factors, such as diet and host genotype) and microbial species interactions (biotic factors, such as competition and mutualism). In contrast, stochastic processes include dispersal and ecological drift (random fluctuations in species relative abundances). Microbial community assembly has been mainly studied from a deterministic perspective, where findings have shown clear impacts of environmental and host factors on gut microbial composition. Until recently, stochastic processes received relatively little attention, and few studies examined the relative importance of each type of processes. Next I will review our current understanding of important factors in shaping gut microbiota, and discuss them in the context of ecological processes described above.

Diet

Diet is one of the most important deterministic factors shaping gut microbiota structure. In mammals, diet appears to be a major driving force, as convergent evolutions of gut microbiota were found in mammals of the same dietary classification (carnivores, omnivores and herbivores) (Ley *et al.*, 2008; Muegge *et al.*, 2011; Delsuc *et al.*, 2014). Likewise, gut microbiota of fruit-feeding *Drosophila* species differ from those of flower-feeding species (Chandler *et al.*, 2011). Similar patterns were observed within the same host species. For example in human, distinct differences in gut microbiota were found in populations on "modern western" and "rural" diet.

The composition of human gut microbiota is thought to be relatively stable during adulthood. One possible explanation for this stability might be the relatively stable long-term dietary behavior. Studies have showed that specific dominant bacterial taxa are associated with macronutrients in the diet, particularly protein, animal fat and plant carbohydrates (David *et al.*, 2015). Diet can rapidly and reproducibly alter gut microbial community structure within a single day (David *et al.*, 2015). It still remains unclear that how the short-term diet shift influence gut microbiota in the long run due to the lack of long term record of host dietary items. One recent study have followed two human subjects for 1 year, and showed that host gut microbiota are relatively stable with minor changes induced by altered diet (David *et al.*, 2014).

Host gut physiology

Host gut physiology is another important driver of gut microbial composition. Physical conditions (such as temperature, morphology, gut peristalsis and so on) obviously have the potential to affect gut microbial composition. For example, microbes with an optimum growth temperature of about 37°C are favored in mammalian gut. Gut morphology is often related to

dietary preferences, while carnivores have simple guts, herbivores evolved complex gut structures suitable for digestion of complex plant polysaccharides by specialized microbes (Ley *et al.*, 2008). Gut morphology can have a large impact on gut microbes. As in the case of giant pandas, despite living on bamboo-based diet, they have simple guts and harbor Carnivore-like gut microbiota similar to their bear relatives. Besides the physical conditions, host-derived substances can also affect gut microbiota. For example, mucin in the intestinal mucus layers can be utilized by some bacteria in the shortage of nutrient sources. Gut immune system, by releasing antimicrobial peptides and antibody IgA into the lumen, can regulate the composition of gut microbial community as well (Brown *et al.*, 2013). Many factors (e.g. host genotype, age, gender) can influence the gut physiology and through which act on the gut microbial communities.

Host genotype

Evidences of host genetic effect on gut microbiota come from the influence of specific genes on gut microbial composition, and the correlation between the overall genetic relatedness of the hosts and microbiota similarity. In mice and human, studies have linked multiple genetic loci including those involved in immune system with shifts in microbial composition (Khachatryan *et al.*, 2008; Salzman *et al.*, 2010; Benson *et al.*, 2010; Rausch *et al.*, 2011; McKnite *et al.*, 2012), providing clear evidence of host genetic control of gut microbes. To date the most comprehensive study in human was conducted on 416 pairs of twins to investigate the correlation between host genetics and gut microbiota (Goodrich *et al.*, 2014). Gut microbiota were more similar between twins than between unrelated individuals, and also more similar between monozygotic twins than between dizygotic twins. This study also found that a few families within phylum Firmicutes were highly heritable.

Dispersal

Dispersal is the process of the movement of an individual (and the taxon it represents) from one location to another. Dispersal tends to homogenize local communities and erase biogeographic patterns created by selection and ecological drift. Bacteria, being small and abundant, have the potential to disperse to great distance, as recapitulated by the hypothesis that "everything is everywhere, but the environment selects" (Becking, 1934). Accordingly, many free-living bacteria are found to be globally distributed, indicating high dispersal rate (Martiny *et al.*, 2011).

Unlike free-living bacteria, members of gut microbiota do not appear to grow outside their host and are most likely transmitted through close host-host interactions. It follows that the dispersal of gut microbiota is rate limited. Each host can be viewed as an island-like patch of habitat occupied by microbial colonists, and the host population as a metacommunity linked through the dispersal of gut microbiota among hosts (Leibold *et al.*, 2004; Vellend, 2010; Costello *et al.*, 2012; Mihaljevic, 2012; Martínez *et al.*, 2015; Christian *et al.*, 2015). With limited dispersal, we would predict a distance-decay relationship where hosts farther apart from each other should exhibit more distinct communities of gut microbes.

Accordingly, recent studies have found higher degree of microbiota similarity among family members than between unrelated individuals (Turnbaugh *et al.*, 2009; Lee *et al.*, 2011; Yatsunenko *et al.*, 2012; Tims *et al.*, 2013). Similarly, comparing gut microbiota among different host populations have revealed interesting biogeographic patterns (Yatsunenko *et al.*, 2012; Linnenbrink *et al.*, 2013; Maurice *et al.*, 2015). However, these results do not necessarily demonstrate dispersal limitation because familial similarities or biogeographic patterns can be caused by closer genetic relationship, shared diet or common environment as well. A study designed to investigate the effect of dispersal limitation should minimize environmental and host

variation among samples (and therefore selection). Adding additional layers of the complexity, dispersal is affected by host lifestyle, social interaction, population density and migration.

Of particular relevance to the study of gut microbiota dispersal is the maternal effect. Traditionally, it is believed that infants are born sterile. However, recent studies show that infant's meconium is non-sterile (Jiménez *et al.*, 2008; Gosalbes *et al.*, 2013), suggesting internal maternal transmission of the initial inoculum. After birth, the guts of newborns were rapidly colonized by maternal microbes from birth canal, breast milk and skin and by microbes in the early environment (O'Toole and Claesson, 2010). A more direct mode of transmission used by many animals is feeding feces from the adult to the baby (coprophagy), thus ensuring the appropriate gut microorganisms colonization. For example, newly hatched termites consume the feces of the adults (Zilber-Rosenberg and Rosenberg, 2008). In the case of koala, the mother feeds the baby at the weaning stage "fecal pap", which contains the bacteria necessary to digest leaves (Osawa *et al.*, 1993). By transmitting gut microbiota to offspring, mother has the potential to influence the phenotype and fitness of offspring in addition to her direct genomic contribution (thereby exerting a maternal effect). Therefore, mother-offspring transmission can contribute to phenotypic plasticity in adaptive evolution and promote co-evolution of the gut microbial community with host species.

Temporal and spatial dynamics of gut microbiota

One major goal of microbial ecology is to understand the how ecological processes generate and maintain diversity across time and space. Here I reviewed current understanding of temporal and spatial dynamics of the gut microbiota in the context of deterministic and stochastic processes that could give rise to such patterns.

Stochastic event such as early colonization play an important role in shaping gut

microbial diversity. Mother makes a large contribution to the initial species pools for colonization (see above). The early colonizers play important role in the proper maturation and development of host immune system and gastrointestinal tract (Mackie *et al.*, 1999; Collado *et al.*, 2012). Inadequate or disrupted postnatal acquisition of infant gut microbiota is associated with an increased risk of immune-mediated diseases, such as allergic rhinitis, asthma, celiac disease, type1 diabetes, and inflammatory bowel disease (Murgas Torrazza and Neu, 2011; Funkhouser and Bordenstein, 2013).

After initial colonization, gut microbiota undergo consecutive ecological successions (Koenig *et al.*, 2011; Lozupone *et al.*, 2012) driven mainly by deterministic factors. With the introduction of solid food and weaning, early colonizers (aerobes and facultative anaerobes) are taken over by strict anaerobes, and infant gut microbiota start to converge towards an adult-like structure (Palmer *et al.*, 2007). It remains unclear whether the "maternal signature" persists to adulthood, and how early colonizer impact later colonization (e.g. prepare the ground for later colonizer or "educate" the immune system to help retain particular microbes). Host undergoes several life stages, during which the diversity (richness) and stability of gut microbiota changes (O'Toole and Claesson, 2010). In human, adult gut microbiota appear to be stable in the absence of disturbance, while combination of deterministic and stochastic events such as disease onset, antibiotic treatment, changes in diet, exposure to new species pool can cause either reversible or irreversible disruptions of gut microbial community (Lozupone *et al.*, 2012; Ursell *et al.*, 2012; Faith *et al.*, 2013; David *et al.*, 2014). Alteration of gut microbiota has been found in elderly people, with decrease in species diversity and Bifidobacteria (Biagi *et al.*, 2010; Claesson *et al.*, 2011). This shift has been associated with changes in host physiology during aging, such as reduction of gut motility and alteration of diet due to reduced dentition (Claesson *et al.*, 2011; 2012).

Besides the general trend described above in the lifetime of the host, gut microbiota can also show temporal patterns on a shorter timescale. Seasonal variation has been observed in gut microbiota of human and wild animals such as wild mice, black howler monkeys, giant pandas, ground squirrels and brown bears (Carey *et al.*, 2013; Davenport *et al.*, 2014; Maurice *et al.*, 2015; Amato *et al.*, 2015; Xue *et al.*, 2015; Sommer *et al.*, 2016). In most cases, it has been suggested that the seasonal variation is likely driven by seasonal variation in diet availability. The latter two hosts represent more extreme cases where hosts undergo hibernation, and gut microbiota differ because of different host physiology between the torpor bout and active phases. To date, most studies of seasonal variation were performed within a year, leaving the question open whether the observed patterns are recurring.

Spatial variation in microbial communities can be driven by environmental factors that vary across space when dispersal is limited. Yatsunenko *et al.* (2012) found that humans occupying different geographic locations have distinct gut microbial community composition. This biogeographic pattern could suggest limited dispersal, but also could be explained by variations in diet and host genetics. A few recent studies tried to tease apart the effects of environmental selection and dispersal limitation. For example, a recent study found evidences suggesting that dispersal limitation likely contribute to the geographic variation in wild mice populations (Linnenbrink *et al.*, 2013). Martínez *et al.* (2015) quantified various ecological processes affecting community assembly and found that dispersal was the dominant process shaping the gut microbiota structure in Papua New Guinea but not in United States. To date, the relative contribution of each process to the spatial variation in microbial community remains largely unknown.

Gut microbiota of wild animal populations

Studying gut microbiota in host's natural environment is important because it is where the actual actions of microbes-host coevolution take place. To date, most gut microbiota studies were performed either solely on human populations or on captive animals in controlled laboratory settings. The ecological forces that shape gut microbial community structure in wild animal populations remain largely unknown. While laboratory animals have been invaluable in dissecting the effect of environmental and host factors on gut microbiota, the settings are unnatural, highly simplified and the gut microbiota of captive animals are often not representative of those in their wild counterparts (Nelson *et al.*, 2013; Amato *et al.*, 2013). It is unclear to what extent findings discovered in laboratory settings can be generalized to gut microbiota of hosts living in natural habitats.

Unlike in controlled settings, wild animals experience temporal and spatial variations in their environment such as climate, habitat, diet availability, population density and social interactions, and their gut microbiota may exhibit temporal and spatial patterns that are not observed in laboratory. Therefore, natural environment provides a rich ecological context to study the underpinning ecological processes and their interactions. For example, black howler monkey species varies their diet based on seasonal and spatial food availability, and exhibit distinct gut microbial communities across seasons and habitats (Amato *et al.*, 2013; 2015). Social interactions can impact the dispersal pattern of gut microbiota among wild animals. Social interactions might vary in response to abiotic factors such as temperature and weather, as well as population density and resource availability. Recent studies on baboons and plateau pikas showed that gut microbiota diversity was correlated with social interaction and population density, and could be explained by dispersal variations across populations (Tung *et al.*, 2015; Li *et al.*, 2016).

Studies of gut microbiota in wild animal populations face a variety of challenges. First of all, we need to find a good system where wild animals live with minimal human contact and intervention, and utilize only natural resources. Researcher can perform non-invasive sample collection on identified host individuals. Secondly, the host population should be well monitored and host ecology and biology are well characterized. As there are literally an unlimited number of measurable variables, it is important for studies to be hypothesis-driven so they can be more focused and productive. Thirdly, it is worth pointing out that wild animal studies offer limited control over the subject and potential confounding factors, and can never achieve the same level of control offered in a laboratory setting. Therefore, it is prudent to have a good study design and obtain sufficient data in order to achieve meaningful findings.

Synopsis of chapters

In this dissertation, I investigated the ecological forces shaping gut microbial communities in three wild animal populations: a baboon population in Africa, a red squirrel population in Canada and *Anolis* lizard species in Puerto Rico, with each tackling the problem from a different angle.

Chapter 2 describes a longitudinal and cross-sectional study of gut microbiota in a well-studied population of wild baboons. The most unique feature of this study is the analysis of longitudinal fecal samples collected over 13 years with detailed metadata of host diet, gender, age, reproductive status and social ranks. We found that baboon gut microbiota were typical of omnivorous primates, and host age and diet had strong effects on gut microbial composition. The most striking finding from this study is that baboon gut microbiota appeared to be highly dynamic such that samples collected from the same individual only a few days apart were as different from each other as samples collected over 10 years apart. Understanding these forces in

wild baboon in their natural habitat provided a valuable comparative context that enriches scientific perspectives on the evolution of human gut microbiota.

In Chapter 3, I conducted a comprehensive study of gut microbiota of a well-studied red squirrel population. Red squirrels defend their exclusive territories over lifetime and thus spend most time in solitude. Therefore red squirrels represent a very attractive system to study the effect of dispersal on gut microbial diversity because unlike human and other wild animals studied so far, red squirrels generally do not move around within the population. For the first time, this study revealed significant spatial patterns of gut microbiota within a host population, suggesting limited dispersal could play a role in shaping and maintaining the structure of gut microbial communities. I found a remarkable seasonal rhythm in red squirrel's gut microbial composition that was clearly driven by seasonal variation in diet availability. Despite the environmental effects, we found evidences for individuality and maternal effect in red squirrel gut microbiota. However, host genetics do not seem to be a significant driver.

Adaptive radiations provide unique opportunities to test whether and how recent ecological and evolutionary diversification of host species structures the composition of microbial communities. In Chapter 4, I examined differences in the gut microbiota of six species of Puerto Rican *Anolis* lizards characterized by the evolution of distinct "ecomorphs" (trunk-crown, trunk-ground, grass-bush), as well as other two species in Florida. Substantial variations in gut microbiota composition were observed within each species and ecomorph. Host phylogeny is weakly correlated with gut microbial composition. Geographic effect was also observed in allopatric conspecifics. Collectively, our results indicate that gut microbial communities are only weakly shaped by the diversification of their lizard hosts due to the strikingly high levels of microbial variation observed within Anolis species.

In summary, I have characterized the temporal, spatial and phylogenetic patterns of gut microbiota in three different wild animal hosts, and identified various ecological factors underlying these patterns. One common conclusion emerged from these studies is that gut microbial diversity in natural environment is strongly associated with diet. Hosts in homogeneous environment harbor stable gut microbiota, while hosts exposed to highly variable environment have highly dynamic gut microbial profiles. These findings may contribute towards a better understanding of how ecological processes govern the gut microbial diversity in natural environment.

References

Amato, K.R., Leigh, S.R., Kent, A., Mackie, R.I., Yeoman, C.J., Stumpf, R.M., et al. (2015) The Gut Microbiota Appears to Compensate for Seasonal Diet Variation in the Wild Black Howler Monkey (Alouatta pigra). *Microb Ecol* **69**: 434–443.

Amato, K.R., Yeoman, C.J., Kent, A., Righini, N., Carbonero, F., Estrada, A., et al. (2013) Habitat degradation impacts black howler monkey (Alouatta pigra) gastrointestinal microbiomes. *The ISME Journal* **7**: 1344–1353.

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005) Host-Bacterial Mutualism in the Human Intestine. *Science* **307**: 1915–1920.

Becking, L.B. (1934) Geobiologie of Inleiding Tot de Milieukunde [Geobiology or Introduction to the Science of the Environment].

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., et al. (2010) 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.* **107**: 18933–18938.

Biagi, E., Nylund, L., Candela, M., Ostan, R., Bucci, L., Pini, E., et al. (2010) Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS ONE* **5**: e10667.

Bik, E.M., Long, C.D., Armitage, G.C., Loomer, P., Emerson, J., Mongodin, E.F., et al. (2010) Bacterial diversity in the oral cavity of 10 healthy individuals. *The ISME Journal* **4**: 962–974.

Brown, E.M., Sadarangani, M., and Finlay, B.B. (2013) The role of the immune system in governing host-microbe interactions in the intestine. *Nat. Immunol.* **14**: 660–667.

Carey, H.V., Walters, W.A., and Knight, R. (2013) Seasonal restructuring of the ground squirrel gut microbiota over the annual hibernation cycle. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **304**: R33–42.

Chandler, J.A., Lang, J.M., Bhatnagar, S., Eisen, J.A., and Kopp, A. (2011) Bacterial Communities of Diverse Drosophila Species: Ecological Context of a Host–Microbe Model System. *PLOS Genet* **7**: e1002272.

Cho, I. and Blaser, M.J. (2012) The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* **13**: 260–270.

Christian, N., Whitaker, B.K., and Clay, K. (2015) Microbiomes: unifying animal and plant systems through the lens of community ecology theory. *Front Microbiol* **6**: 869.

Claesson, M.J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., de Weerd, H., Flannery, E., et al. (2011) Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc. Natl. Acad. Sci. U.S.A.* **108 Suppl 1**: 4586–4591.

Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.M., Cusack, S., et al. (2012) Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**: 178–184.

Claesson, M.J., O'Sullivan, O., Wang, Q., Nikkilä, J., Marchesi, J.R., Smidt, H., et al. (2009) Comparative Analysis of Pyrosequencing and a Phylogenetic Microarray for Exploring Microbial Community Structures in the Human Distal Intestine. *PLoS ONE* **4**: e6669.

Collado, M.C., Cernada, M., Baüerl, C., Vento, M., and Pérez-Martínez, G. (2012) Microbial ecology and host-microbiota interactions during early life stages. *Gut Microbes* **3**: 352–365.

Costello, E.K., Stagaman, K., Dethlefsen, L., Bohannan, B.J.M., and Relman, D.A. (2012) The application of ecological theory toward an understanding of the human microbiome. *Science* **336**: 1255–1262.

Davenport, E.R., Mizrahi-Man, O., Michelini, K., Barreiro, L.B., Ober, C., and Gilad, Y. (2014) Seasonal variation in human gut microbiome composition. *PLoS ONE* **9**: e90731.

David, L.A., Materna, A.C., Friedman, J., Campos-Baptista, M.I., Blackburn, M.C., Perrotta, A., et al. (2014) Host lifestyle affects human microbiota on daily timescales. *Genome Biology* **15**: R89.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., et al. (2015) Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**: 559–563.

Delsuc, F., Metcalf, J.L., Wegener Parfrey, L., Song, S.J., Gonzalez, A., and Knight, R. (2014) Convergence of gut microbiomes in myrmecophagous mammals. *Mol. Ecol.* **23**: 1301–1317.

Drasar, B.S. (1967) Cultivation of anaerobic intestinal bacteria. *The Journal of Pathology and Bacteriology* **94**: 417–427.

Escherich, T. (1885) Die Darmbacterien des Neugeborenen und Säuglings.

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., et al. (2013) The long-term stability of the human gut microbiota. *Science* **341**: 1237439.

Frank, D.N., St Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007)

Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U.S.A.* **104**: 13780–13785.

Funkhouser, L.J. and Bordenstein, S.R. (2013) Mom Knows Best: The Universality of Maternal Microbial Transmission. *PLoS Biol* **11**: e1001631–9.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., et al. (2014) Human Genetics Shape the Gut Microbiome. *Cell* **159**: 789–799.

Gosalbes, M.J., Llop, S., Vallès, Y., Moya, A., Ballester, F., and Francino, M.P. (2013) Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clin. Exp. Allergy* **43**: 198–211.

Graessler, J., Qin, Y., Zhong, H., Zhang, J., Licinio, J., Wong, M.-L., et al. (2013) Metagenomic sequencing of the human gut microbiome before and after bariatric surgery in obese patients with type 2 diabetes: correlation with inflammatory and metabolic parameters. *Pharmacogenomics J.* **13**: 514–522.

Hollister, E.B., Gao, C., and Versalovic, J. (2014) Compositional and functional features of the gastrointestinal microbiome and their effects on human health. *Gastroenterology* **146**: 1449–1458.

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

Jiménez, E., Marín, M.L., Martín, R., Odriozola, J.M., Olivares, M., Xaus, J., et al. (2008) Is meconium from healthy newborns actually sterile? *Res. Microbiol.* **159**: 187–193.

Khachatryan, Z.A., Ktsoyan, Z.A., Manukyan, G.P., Kelly, D., Ghazaryan, K.A., and Aminov, R.I. (2008) Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS ONE* **3**: e3064.

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J.I., Knight, R., et al. (2011) Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl. Acad. Sci. U.S.A.* **108 Suppl 1**: 4578–4585.

Koeppel, A., Perry, E.B., Sikorski, J., Krizanc, D., Warner, A., Ward, D.M., et al. (2008) Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc. Natl. Acad. Sci. U.S.A.* **105**: 2504–2509.

Lee, S., Sung, J., Lee, J., and Ko, G. (2011) Comparison of the gut microbiotas of healthy adult twins living in South Korea and the United States. *Appl. Environ. Microbiol.* **77**: 7433–7437.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., et al. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters* **7**: 601–613.

Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005) Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 11070–11075.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., and Gordon, J.I. (2008) Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology* **6**: 776–788.

Ley, R.E., Peterson, D.A., and Gordon, J.I. (2006) Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. *Cell* **124**: 837–848.

Li, H., Qu, J., Li, T., Li, J., Lin, Q., and Li, X. (2016) Pika Population Density Is Associated with the Composition and Diversity of Gut Microbiota. *Front Microbiol* **7**: 425.

Linnenbrink, M., Wang, J., Hardouin, E.A., Künzel, S., Metzler, D., and Baines, J.F. (2013) The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Mol. Ecol.* **22**: 1904–1916.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012) Diversity,

stability and resilience of the human gut microbiota. *Nature* **489**: 220–230.

Mackie, R.I., Sghir, A., and Gaskins, H.R. (1999) Developmental microbial ecology of the neonatal gastrointestinal tract. *Am. J. Clin. Nutr.* **69**: 1035S–1045S.

Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D., and Horner-Devine, M.C. (2011) Drivers of bacterial beta-diversity depend on spatial scale. *Proc. Natl. Acad. Sci. U.S.A.* **108**: 7850–7854.

Martínez, I., Stegen, J.C., Maldonado-Gómez, M.X., Eren, A.M., Siba, P.M., Greenhill, A.R., and Walter, J. (2015) The Gut Microbiota of Rural Papua New Guineans: Composition, Diversity Patterns, and Ecological Processes. *Cell Rep* **11**: 527–538.

Maurice, C.F., Knowles, S.C.L., Ladau, J., Pollard, K.S., Fenton, A., Pedersen, A.B., and Turnbaugh, P.J. (2015) Marked seasonal variation in the wild mouse gut microbiota. *The ISME Journal* **9**: 2423–2434.

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A., et al. (2012) Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PLoS ONE* **7**: e39191.

Mihaljevic, J.R. (2012) Linking metacommunity theory and symbiont evolutionary ecology. *Trends Ecol. Evol. (Amst.)* **27**: 323–329.

Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., et al. (2012) Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biology* **13**: R79.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., Gonzalez, A., Fontana, L., et al. (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**: 970–974.

Murgas Torrazza, R. and Neu, J. (2011) The developing intestinal microbiome and its

relationship to health and disease in the neonate. *J Perinatol* **31 Suppl 1**: S29–34.

Naseribafrouei, A., Hestad, K., Avershina, E., Sekelja, M., Linløkken, A., Wilson, R., and Rudi, K. (2014) Correlation between the human fecal microbiota and depression. *Neurogastroenterol. Motil.* **26**: 1155–1162.

Nelson, T.M., Rogers, T.L., Carlini, A.R., and Brown, M.V. (2013) Diet and phylogeny shape the gut microbiota of Antarctic seals: a comparison of wild and captive animals. *Environ Microbiol* **15**: 1132–1145.

Osawa, R., Blanshard, W.H., and Ocallaghan, P.G. (1993) Microbiological Studies of the Intestinal Microflora of the Koala, Phascolarctos-Cinereus .2. Pap, a Special Maternal Feces Consumed by Juvenile Koalas. *Australian Journal of Zoology* **41**: 611–620.

O'Toole, P.W. and Claesson, M.J. (2010) Gut microbiota: Changes throughout the lifespan from infancy to elderly. *International Dairy Journal* **20**: 281–291.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007) Development of the Human Infant Intestinal Microbiota. *PLoS Biol* **5**: e177.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., et al. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**: 59–65.

Rausch, P., Rehman, A., Künzel, S., Häsler, R., Ott, S.J., Schreiber, S., et al. (2011) Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proc. Natl. Acad. Sci. U.S.A.* **108**: 19030–19035.

Salzman, N.H., Hung, K., Haribhai, D., Chu, H., Karlsson-Sjöberg, J., Amir, E., et al. (2010) Enteric defensins are essential regulators of intestinal microbial ecology. *Nat. Immunol.* **11**: 76–83.

Savage, D.C. (2001) Microbial Biota of the Human Intestine: A Tribute to Some Pioneering Scientists. *Curr. Issues Intest. Microbiol* **2**: 1–15.

Schaedler, R.W., Dubos, R., and Costello, R. (1965) The Development of the Bacterial Flora in the Gastrointestinal Tract of Mice. *J Exp Med* **122**: 59–66.

Sommer, F. and Bäckhed, F. (2013) The gut microbiota-masters of host development and physiology. *Nature Reviews Microbiology* **11**: 227–238.

Sommer, F., Ståhlman, M., Ilkayeva, O., Arnemo, J.M., Kindberg, J., Josefsson, J., et al. (2016) The Gut Microbiota Modulates Energy Metabolism in the Hibernating Brown Bear Ursus arctos. *Cell Rep* **14**: 1655–1661.

Tims, S., Derom, C., Jonkers, D.M., Vlietinck, R., Saris, W.H., Kleerebezem, M., et al. (2013) Microbiota conservation and BMI signatures in adult monozygotic twins. *The ISME Journal* **7**: 707–717.

Tung, J., Barreiro, L.B., Burns, M.B., Grenier, J.-C., Lynch, J., Grieneisen, L.E., et al. (2015) Social networks predict gut microbiome composition in wild baboons. *eLife Sciences* **4**: e1002358.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., et al. (2009) A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484.

Ursell, L.K., Metcalf, J.L., Parfrey, L.W., and Knight, R. (2012) Defining the human microbiome. *Nutrition Reviews* **70**: S38–S44.

Vellend, M. (2010) Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology* **85**: 183–206.

Vrieze, A., Van Nood, E., Holleman, F., Salojärvi, J., Kootte, R.S., Bartelsman, J.F.W.M., et al. (2012) Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology* **143**: 913–6.e7.

Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., Levison, B.S., Dugar, B., et al. (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**: 57–63.

Xue, Z., Zhang, W., Wang, L., Hou, R., Zhang, M., Fei, L., et al. (2015) The bamboo-eating giant panda harbors a carnivore-like gut microbiota, with excessive seasonal variations. *mBio* **6**: e00022–15.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., et al. (2012) Human gut microbiome viewed across age and geography. *Nature* **486**: 222–227.

Zilber-Rosenberg, I. and Rosenberg, E. (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews* **32**: 723–735.

Chapter 2. Development, diet and dynamism: longitudinal and cross-sectional predictors of gut microbial communities in wild baboons

Formatted as a co-authored manuscript and published as:

Ren T, Grieneisen LE, Alberts SC, Archie EA, Wu M. (2015) Development, diet and dynamism: longitudinal and cross-sectional predictors of gut microbial communities in wild baboons. *Environmental Microbiology*, **18**(5): 1312–1325.

DOI: 10.1111/1462-2920.12852

Abstract

Gut bacterial communities play essential roles in host biology, but to date we lack information on the forces that shape gut microbiota between hosts and over time in natural populations. Understanding these forces in wild primates provides a valuable comparative context that enriches scientific perspectives on human gut microbiota. To this end, we tested predictors of gut microbial composition in a well-studied population of wild baboons. Using cross-sectional and longitudinal samples collected over 13 years, we found that baboons harbor gut microbiota typical of other omnivorous primates, albeit with an especially high abundance of *Bifidobacterium*. Similar to previous work in humans and other primates, we found strong effects of both developmental transitions and diet on gut microbial composition. Strikingly, baboon gut microbiota appeared to be highly dynamic such that samples collected from the same individual only a few days apart were as different from each other as samples collected over 10 years apart. Despite the dynamic nature of baboon gut microbiota, we identified a set of core taxa that is common among primates, supporting the hypothesis that microbiota codiversify with their host species. Our analysis identified two tentative enterotypes in adult baboons that differ from those of humans and chimpanzees.

Introduction

Vertebrate gut microbiota play important roles in host biology, including immune regulation, energy acquisition, vitamin synthesis, and disease risk (Turnbaugh et al., 2006; Hooper et al., 2012; Bengmark, 2013; Morgan et al., 2013). There is mounting evidence that variation in the functions of gut microbiota are mediated by variation in gut microbial composition both within and between hosts (Turnbaugh et al., 2009; Greenblum et al., 2011; Hooper et al., 2012; Bengmark, 2013; Iida et al., 2013; Karlsson et al., 2013; Koeth et al., 2013; Markle et al., 2013; Viaud et al., 2013). However to date, most such evidence comes from research on humans and captive animal models, leaving large gaps in our understanding of the forces that shape gut microbial composition in wild vertebrates, both between individuals and within the same individual over time. Filling these gaps is especially important for wild primates, in part because such information helps reveal which of the forces that shape human gut microbiota are common across primates, and which are unique to humans and perhaps a consequence of modern human lifestyles (Yildirim et al., 2010; Degnan et al., 2012; Amato, 2013; Moeller et al., 2014).

To help address these gaps, we used cross-sectional and longitudinal sampling to characterize distal gut bacterial communities over an unusually long, 13-year time span in a well-studied population of wild baboons. Baboons provide an especially relevant comparative system for understanding variation in human gut microbiota because of their relatively close evolutionary relationship to humans, and because baboons lead a terrestrial, savannah-dwelling lifestyle that is thought to resemble the ecology of early humans (DeVore and Washburn, 1963; Codron et al., 2008; Sponheimer et al., 2013). Specifically, we worked with the Amboseli Baboon Research Project in Kenya (Alberts and Altmann, 2012), where longitudinal, individual-

based research on demography, developmental milestones, social relationships, diet, and climate provided an especially rich context within which to understand individual-level variation in the structure of gut microbial communities. Our main objectives were to: (i) characterize the basic structure of gut microbiota in wild baboons, (ii) gain a multivariate understanding of the relative importance of development, social relationships, diet, and climate in predicting gut microbial community structure, (iii) understand patterns of longitudinal change in baboon gut microbial communities, and (iv) test whether gut microbiota in wild baboons contain a core set of microbial taxa and enterotypes. Throughout, we discuss our results in the context of what is known about human and other primate gut microbiota.

Results and discussion

General patterns in the baboon gut microbial profile

We analyzed distal gut microbial composition using 107 fecal samples from 24 baboons collected between 1994 and 2009 (Table 1). For 13 baboons, we analyzed multiple samples (range = 3 to 10 samples per baboon; time span between longitudinal samples ranged from 2 days to 13 years). From these 107 samples, we generated 358,428 high-quality 16S rRNA reads, yielding an average 3,350 reads per sample and 7,201 total OTUs using a 97% identity cutoff. This dataset was rarefied to 1,500 (n = 107 samples from 24 baboons; Table 1) and 3,000 reads per sample (n = 54 samples from 17 baboons; Table S1). The 1,500- and 3,000-read datasets were highly similar in microbial composition at the OTU level (Pearson correlation coefficient = 0.983, $P = 0.001$). Furthermore, Mantel tests correlating compositional dissimilarity in samples rarefied to 1,500 versus 3,000 reads revealed a high level of congruence between these datasets (Mantel tests for Bray-Curtis dissimilarities: $r = 0.996$, $P = 0.001$; unweighted UniFrac: $r = 0.978$, $P = 0.001$; weighted UniFrac: $r = 0.996$, $P = 0.001$). As a result, the results we present in

the main text rely on the 1,500-read dataset; where appropriate, we repeat analyses using the 3,000-read dataset and present them as Supplementary Information.

Taxonomic assignment revealed representatives from 11 bacterial phyla and 90 bacterial genera. Similar to other mammalian gut microbial communities, the four most common phyla included Firmicutes (48.8% of reads), Actinobacteria (17.2%), Bacteroidetes (7.2%), and Proteobacteria (4.1%). However, compared to humans and other primates (Ley et al., 2008; Yildirim et al., 2010), samples from wild baboons harbored a much higher percentage of Actinobacteria, of which 97.8% were assigned to the genus *Bifidobacterium*, and a relatively smaller proportion of Bacterioidetes (Figure 1). *Bifidobacterium* is dominant in the gut flora of breastfed human infants (Turroni et al., 2012), where it is thought to play a role in the digestion of the complex carbohydrates in human milk. Accordingly, in humans, the percentage of *Bifidobacterium* decreases dramatically with age and it comprises only 3-6% of the adult gut flora. In baboons, grasses and other fiber-rich foods were common in the diet, and *Bifidobacterium spp*. may be important in digesting the high fiber content of these foods.

Notably, nearly one fifth (20.8%) of reads were unclassified and potentially novel at the phylum level. These unclassified OTUs were unlikely to be sequencing artifacts because the vast majority (86.1%) appeared in more than one sample, and their distribution among samples was indistinguishable from that of classified OTUs (Figure S1). Moreover, 54.0% of these unclassified OTUs were ≥ 90% identical to OTUs found in other mammalian fecal samples (Ley et al., 2008).

Previous studies have found that mammalian gut microbial composition is strongly associated with host diet and phylogeny (Ley et al., 2008; Yildirim et al., 2010; Hong et al., 2011; Degnan et al., 2012; Bolnick et al., 2014; Delsuc et al., 2014). Therefore, we compared the gut microbiota of our baboon samples to those of other mammals. As expected, the 13 of 14

fecal samples (one from each adult baboon) from our study clustered with other primates, especially those with omnivorous diets (Figure 2). Samples from within the same order or diet group were significantly more similar than samples from different orders or diet groups (Table S2). Hence, host phylogeny and diet both seem to play dominant roles in determining variation in gut microbial composition between host species. We note that these patterns were based on the most abundant bacteria in each sample because of the small number of reads in the mammal dataset (Ley et al., 2008).

Juvenile baboons exhibited lower bacterial alpha diversity, but higher variance than than adults

Alpha diversity is an important component of microbial diversity in the gut, especially in the context of microbiota development and pathogen resistance (Dillon et al., 2005; McKenna et al., 2008; Degnan et al., 2012; Flores et al., 2012; Yatsunenko et al., 2012; Ahn et al., 2013). Furthermore, given that some of our samples were collected more than 15 years prior to analysis (Table 1), we were concerned that DNA degradation might affect microbial alpha diversity in our samples, and hence our ability to characterize gut microbial composition. However, we found no evidence that older samples exhibited lower alpha diversity than younger samples (linear mixed models with sample age in years as a fixed effect and host identity as a random effect: species richness: β = -1.37, P = 0.76; Shannon's H: β = 0.004, P = 0.89; Chao1: β = -4.14, P = 0.63; Faith's phylogenetic diversity: β = -0.19, P = 0.28).

In addition to sample age, we tested several other predictors of gut microbial alpha diversity, including host age, sex, rainfall in the 30 days prior to sample collection, current social group, natal social group, current social group size, host diet composition, host diet alpha diversity, adult social rank, and for adult females only, reproductive state (as pregnant, lactating,

or ovarian cycling). Because host identity significantly predicted variation in alpha diversity for 2 of the 4 measures (ANOVA; species richness: $F_{(12,83)} = 2.03$, $P = 0.031$; Shannon's H: $F_{(12,83)} = 2.78$, $P = 0.003$; Chao1: $F_{(12,83)} = 1.39$, $P = 0.187$; Faith's phylogenetic diversity: $F_{(12,83)} = 1.72$, $P = 0.078$), individual identity was included as a random effect in all linear mixed models. In prior studies on humans and chimpanzees, age was a primary predictor of gut microbial alpha diversity (Degnan et al., 2012; Yatsunenko et al., 2012). However, in our data set, neither age nor any other fixed effects predicted any of the four measures of alpha diversity in linear mixed models. However, when we divided samples into juveniles and adults rather than testing age as a continuous variable, we found that adults had greater species richness (Wilcoxon rank-sum test; W = 1723.5, $P = 0.049$) and Shannon's H (Wilcoxon rank-sum test; W = 1783, $P = 0.019$) than juveniles. Additionally, similar to one prior study in humans (Yatsunenko et al., 2012), we found that infants and juveniles exhibited significantly higher variance in Shannon's H than adult baboons (Figure 3; Brown-Forsyth test $F_{(1, 104.876)} = 6.988$, $P = 0.009$). Taken together, these results indicate that gut microbiota may be less diverse and less stable in young baboons as compared to adults, suggesting that the transition to adulthood marks a developmental milestone in the microbiota.

Variation in microbial composition was best explained by host age, diet, and rainfall

We tested whether several host traits and environmental factors were associated with variation in baboon gut microbial composition, including host identity, host age, host sex, rainfall in the 30 days prior to sample collection, current social group, natal social group, and group size. To test these factors, we first performed an exploratory principal coordinates analysis (PCoA) on weighted UniFrac dissimilarities. Overall, 37.5% of the global variation was explained by the first three principal coordinates (PC1 = 21%, PC2 = 9.3%, PC3 = 7.2%). Visual

inspection revealed no obvious clustering patterns by any of our predictor variables (Figure S2 A-E).

To further test which factors best explained variation in baboon gut microbial composition, we carried out canonical correspondence analysis (CCA) (Palmer, 1993) at the phylum, genus, and OTU levels (Table 2). For the main dataset (n = 107 samples), we again tested the effects of host identity, host age, sex, rainfall, and aspects of social group membership. Interestingly, no factors explained significant variation at the OTU level, perhaps because closely related species are often ecologically interchangeable (Harvey and Pagel, 1991), and ecological patterns in bacterial communities may be more apparent at higher taxonomic levels. In support, some recent studies have found ecological coherence among higher bacterial taxonomic ranks (Fierer et al., 2007; Lozupone and Knight, 2007; von Mering et al., 2007; Fulthorpe et al., 2008; Pointing et al., 2009; Philippot et al., 2010; Koeppel and Wu, 2012). Indeed, we were able to explain significant shifts in high-level taxa associated with changes in environment. Specifically, rainfall and age predicted significant variation at both the phylum level (Table 2: rainfall, P = 0.02; age, P = 0.02) and the genus level (rainfall, P = 0.01; age, P = 0.02). Interestingly, while age predicted beta diversity, when baboon infants, juveniles, and adults were considered separately, age was no longer significant, suggesting that either the subsets of samples do not have enough statistical power or developmental transitions are more important than age *per se* (Table 2). Host sex was significant when we considered infants alone, perhaps due to differences in maternal care as a function of infant sex (Nguyen et al. 2012). These sex-differences seem to disappear in adulthood, however.

While some prior studies have found evidence for social group membership on gut microbial composition (Degnan et al., 2012; Yatsunenko et al., 2012), including in our own population (Tung et al. in press), the wide temporal distribution of samples in our data set

probably made it difficult to detect such effects. We found no other physiological or social effects on gut microbial composition, including female reproductive state, social group size, or male or female dominance rank.

The effects of rainfall on microbiota may be linked to seasonal changes in either diet or drinking water availability. To test the specific effects of diet, we conducted a second CCA using only the subset of 76 individuals (excluding infants) for which we had data on the time spent foraging on different food types. In this new model, we found several effects of diet (Table 2; Table S4). First, diet alpha diversity (Shannon's H for diet components) explained significant variation at both the phylum (P =0.02) and genus levels (P =0.05), while diet richness (the number of distinct food types; Table S3) did not, suggesting that dietary evenness rather than a high number of dietary components is important to the gut microbial composition. This pattern runs counter to that seen in (Bolnick et al., 2014), which found that dietary richness rather than evenness predicted gut microbial diversity in fish. Second, the dietary tradeoff between the proportion of time spent consuming grass versus fruit in the diet (diet PC1; see Experimental Procedures; Figures S3) was significantly associated with microbial composition at both phylum level (P =0.02) and genus level (P =0.02). .

To assess the influence of sequencing depth on our results, we repeated the CCA using the smaller dataset rarified to 3,000 reads (n = 54 samples; diet information on n = 38 samples). We obtained best models similar to those of 1,500-read dataset, although none of the factors were significant, probably as a result of a loss of statistical power (Table S4).

Finally, to identify which of the four most common bacterial phyla were associated with differences in host age, rainfall, and diet, we performed generalized linear mixed models with a Poisson-link and host identity as a random effect. We found that samples collected in rainier periods harbored a higher proportion of Firmicutes, but less Actinobacteria than samples from

drier months (Table 3). In terms of host age (measured as a continuous variable), younger animals harbored relatively more Actinobacteria, Bacteroidetes, and Proteobacteria but less Firmicutes than older animals, perhaps due to differences in milk consumption or disease susceptibility in animals of different ages. For the subset of samples with diet information, gut microbiota from groups that consumed relatively more fruit and less grass harbored higher levels of Actinobacteria and Proteobacteria and lower levels of Firmicutes and Bacteroidetes than groups consuming low fruit (diet PC1, Table S5). Furthermore, the addition of rainfall significantly improved models with diet factors, indicating that the effects of rainfall are not solely driven by seasonal changes in diet. During the dry season, the baboons drink from small, highly concentrated and qualitatively dirty water holes whereas during rainy months they obtain most of their water from seemingly cleaner, transient rain puddles, which may have consequences for gut microbiota.

Longitudinal sampling reveals that baboon gut microbiota are highly dynamic

Prior research on humans and chimpanzees has found that individuals contain distinct gut microbiota, and that samples from the same individual, even those collected over a year apart, are more similar to each other than they are to samples collected from different hosts over the same time period (Turnbaugh et al., 2009; Caporaso et al., 2011; Degnan et al., 2012; David et al., 2014). However, we found no evidence for such effects in our study subjects. For instance, in the CCA analyses described above, we never observed a significant effect of individual identity at any taxonomic level. Similarly, samples from the same individual were as different from each other as they were from samples collected from different individuals in the same developmental stage (mean ± SE weighted UniFrac dissimilarity: between samples from the same individual = 0.342 ± 0.008; between samples from different individuals at the same stage = 0.345 ± 0.002, *P*

= 0.35). However, we note the sequence depth in our study is lower than those of the previous studies.

This high degree of dynamism in baboon gut microbiota can be visualized by plotting pairwise weighted UniFrac dissimilarities between samples of the same individual as a function of time between sampling points (Figure 4). Microbial communities sampled from the same individual a few days apart were almost as different from each other as samples collected several years apart. Only one of 13 individuals with >3 samples displayed a significant relationship between sampling time interval and microbiota dissimilarity (Table S6). It is unclear why baboon gut microbiota appeared to be so dynamic. One possible explanation is that seasonal variation in the baboons' diets selects for different gut microbial compositions at different times of year, as the availability of fruits, seeds, and vegetation fluctuates with seasonal patterns in plant reproduction. However, such seasonal variation is unlikely to explain turnover on the scale of days or weeks during which baboon diets are more consistent. Another explanation is that wild baboons live in microbially heterogeneous environments, regularly walking through fecal deposits of other species, drinking from waterholes that contain fecal material from livestock and wild mammals, and pulling plants from the ground with their mouths. This could lead to higher turnover in gut microbial species.

Core gut microbiota

The high degree of inter- and intra-individual variation in baboon microbiota raises the question of whether baboon gut microbiota contain a set of core microbial taxa, as is observed in humans (Tap et al., 2009; Martínez et al., 2013). We defined core taxa as taxa present in more than 90% of our 107 samples, assigned at the lowest possible taxonomic level. Despite the dynamic nature of baboon gut microbiota, we found evidence for some core taxa: three at the

family level (Lachnospiraceae, Peptostreptococcaceae, and Veillonellaceae) and four at the genus level (*Faecalibacterium*, *Prevotella*, *Bifidobacterium*, and *Oscillibacter*).

To investigate how core microbiota have changed with host phylogeny, we attempted to identify core gut microbial members of the 57 mammalian species (89 individuals) used in (Ley et al., 2008). Given the large variation in mammalian genomes, diets, and lifestyles, it is not surprising that we did not find any core taxon below the phylum-level that are shared by all mammals. However, when we limited our scope to primates alone, we found two family-level (Ruminococcaceae and Lachnospiraceae) and one genus-level (*Prevotella*) core taxa. Since these taxa are present in most primates surveyed, these core taxa were most likely present in the last common ancestor of primates, suggesting they might be important in the codiversification of the gut microbiota and the primate hosts.

Enterotypes in baboons

Previous studies reported that humans and chimpanzees harbor compositionally similar gut enterotypes (Arumugam et al., 2011; Moeller et al., 2012). To test for the presence of enterotypes in our subjects, we clustered gut microbiota for the 47 (of 107) samples collected from sexually mature, adult baboons by applying the partitioning around means (PAM) clustering method on the Bray-Curtis dissimilarities calculated using genus level abundances (Arumugam et al., 2011). Our analysis revealed an optimum of two clusters (Figure 5. CH index: 39; average silhouette coefficient 0.265; prediction strength: 0.79). Although the silhouette coefficient is comparable to those reported in earlier enterotype studies (Arumugam et al., 2011; Moeller et al., 2012), it would be considered low according to the thresholds proposed more recently (Koren et al., 2013). Therefore, the enterotypes identified in this study are tentative. The genera contributed most significantly to each cluster were *Bifidobacterium*, *Butyrivibrio*,

Megasphaera, and *Olsenella* in enterotype 1 (n=9) and *Oscillibacter* and *Ruminococcus* in enterotype 2 (n=38). The relative abundances of genera in the adult samples are listed in Table S7. These two enterotypes differ from those of humans and chimpanzees. One parsimonious explanation is that enterotypes in humans and chimpanzees may have evolved since the split between apes and old world monkeys ~30 million years ago.

Previous studies found that enterotypes can be replaced within one year in chimpanzees (Moeller et al., 2012) and within one week in wild mice housed in captivity (Wang et al., 2014). We observed enterotype replacements for most baboons when we assessed the samples of the same individual at multiple time points (Figure 6). Enterotypes changed rapidly in baboons, sometimes switching in as little as 45 days. Past studies have suggested that proportion of protein versus carbohydrates in host diet is linked to the host's enterotype (Wu et al., 2011; Wang et al., 2014). However, the baboon enterotypes we found were not significantly associated with any factors tested, including diet diversity (richness, Shannon's H and PCoA axis), age, rainfall, host identity, season, sex or social group (Wilcoxon rank sum test or Fisher's Exact Test). Consistent with the finding of (Wang et al., 2014), we found no enterotypes at the OTU level.

Experimental Procedures

Study subjects and predictors of microbiota structure

Study subjects were wild baboons living in the Amboseli Ecosystem in Kenya, a semi-arid savannah located northeast of Mt. Kilimanjaro (2°40′S, 37°15′E, 1100 m altitude). Since 1971, the baboons in this area have been studied by the Amboseli Baboon Research Project (ABRP) (Alberts and Altmann, 2012). Several types of data are collected throughout the year on known individuals by full-time, experienced observers, 2-3 times per week per group. Here we describe data collection on the specific predictor variables we tested; sample sizes vary

somewhat for each predictor variable because some data were only available for or relevant to some individuals and samples.

Sex, age, and developmental stage. Baboons are sexually dimorphic, and sex is known from conspicuous external genital morphology. Ages were known to within a few days for 20 of 24 animals in our main dataset (n = 107). The remaining 4 individuals immigrated into the population after birth and their ages were estimated using well-defined metrics and comparison to known-age animals (Alberts and Altmann, 1995). These 4 animals had birth dates estimated to be accurate within one year (n = 1), two years (n = 2), or three years (n = 1). As baboons mature, they pass through several developmental stages that may also influence the gut microbiota, including: (i) infancy, during which diet includes both milk and foods from the environment (birth to 1.5 years; n = 22 fecal samples from 9 individuals); (ii) the juvenile period, which begins post weaning (~1.5 years) and ends at sexual maturity (~4.5 years for females; ~5.4 years for males; 38 samples from 10 individuals (Onyango et al., 2013)), and (iii) adulthood, defined by the onset of sexual maturity (n = 47 samples from 14 individuals).

Diet. In addition to the dietary changes associated with the transition from the infant to the juvenile stage, we tested the effect of diet composition on gut microbiota. Specifically, for a subset of subjects (n = 76 fecal samples from 10 juveniles and 14 adults), we estimated diet composition using behavioral sampling on all the juvenile and adult female members of the social group in the 30 days prior to sample collection. Social group members consume similar foods in roughly similar proportions; hence group-level diets provide suitable estimates of the composition of individual diets. The baboons' diets included 11 food categories: (1) grass, including corms, blades, and grass seed heads, (2) gum from the bark of *Acacia xanthophloea*, (3) leaves from herbaceous plants or trees, (4) fruits, (5) blossoms, (6) bark from *A. xanthophloea*, (7) fresh, green seed pods from *Acacia* spp., (8) dried seeds from *Acacia* spp., (9)

invertebrates, (10) liquid from or items in or under dung, and (11) unknown unidentifiable items (Table S3).

We used these data to characterize diet alpha and beta diversity, noting that time spent feeding is not always proportional to the amount of food ingested. Diet alpha diversity was measured as both the total number of foods (diet richness) and dietary Shannon's H using the vegan package in R (Oksanen et al., 2012). Diet beta diversity was estimated via PCoA on a Bray-Curtis dissimilarity matrix of diet composition using vegan. The first three axes of the PCoA explained 80% of the variation in diet (PC1 = 46%, PC2 = 23%, PC3 = 11%); PC1 was associated with a tradeoff in relative proportions of grass (-) versus fruit (+); PC2 was associated with the proportion of invertebrates (-) versus fruit (+); and PC3 was associated with the proportion of the diet attributed to the 'unknown' category (-) (Figures S3-5).

Rainfall. Semi-arid savannah ecosystems are characterized by highly seasonal patterns of rain that may affect diet as well as bacterial exposures through sources of drinking water. Each year, Amboseli experiences a five-month dry season (June – October) during which no rain falls. In the remaining seven months (November - May), the ecosystem receives highly variable amounts of rain (yearly average = 350 mm; range = 141-757 mm) (Alberts et al., 2005). The effects of rainfall were assessed by summing the total amount of rain that fell in the 30 days prior to sample collection.

Social relationships. One prior study has linked aspects of primate social group membership to microbial composition (Degnan et al., 2012). We tested three aspects of social group: (i) the identity of the animal's social group on the day of sample collection; (ii) the size of the animal's social group on the day of sample collection, as the number of members, and (iii) the identity of the animal's natal social group, if known (in baboons, males are the dispersing sex and the current group of an adult male invariably differs from his natal group).

In addition, in baboons, dominance rank has been linked to physiology and health (Sapolsky and Altmann, 1991; Alberts et al., 1992; Gesquiere et al., 2011); hence we also tested for associations between microbial composition and dominance rank. Rank was assigned monthly by observing dyadic agonistic interactions and assigning winners and losers based on the outcome. These wins and losses were used to construct dominance matrices, resulting in an ordinal rank for each member of the group (Hausfater, 1975).

Adult female reproductive status. Prior research has shown that reproductive cycle changes in human women can influence gut microbial composition (Koren et al., 2012). To test this idea, samples from adult females (24 samples from 9 individuals) were assigned to one of three reproductive states: ovarian cycling, pregnant, or lactating, using previously published and well-defined criteria (Altmann, 1973; Wildt et al., 1977; Shaikh et al., 1982; Beehner et al., 2006; Gesquiere et al., 2007).

Sample collection, DNA extraction, and 16S rRNA sequencing

Gut microbiota were characterized from fecal samples. Samples for this analysis spanned 1994 to 2009 and included 144 samples from 32 individuals. Samples were chosen to provide both cross-sectional and longitudinal information, including multiple samples from a subset of 13 individuals. All fecal samples were collected within a few minutes of defecation, after which the sample was mixed and preserved in 95% ethanol. Samples were stored in an evaporative cooling structure (approximate daily maximum temperature of 25°C) until shipment to the US, where they were stored at -80°C. DNA was extracted from each sample by bead beating and phenol-chloroform extraction. For each DNA extract, the V1-V3 hypervariable regions of the 16S rRNA gene were PCR amplified and pyrosequenced as described previously (Ren et al.,

2013) on a 454 Life Science Genome Sequencer FLX platform (University of Virginia Department of Biology Genome Core Facility).

Sequence processing, quality control, and OTU classification

Sequencing reads were processed using the QIIME pipeline (Caporaso et al., 2010). Each read was assigned to a sample by barcode and then filtered to remove reads with: (i) lengths less than 200 base pairs or greater than 550 base pairs, (ii) average Phred equivalent quality scores less than 25, (iii) improper primer or barcode sequences, or (iv) the presence of ambiguous base calls. Eukaryotic, mitochondrial sequences were removed by BLAST search against the SILVA database (Quast et al., 2013). Chloroplast sequences were removed using the Ribosomal Database Project (RDP) classifier (Wang et al., 2007). Chimeric sequences were identified using UCHIME with the de novo detection algorithm and default parameters (Edgar et al., 2011). Read filtering removed 17.8% of the total reads, with the majority removed as chimeras. The remaining reads were clustered to 97% operational taxonomic units (OTUs) by cdhit (Fu et al., 2012). To further remove potential sequencing artifacts, we excluded any OTU with ≤ five reads across all samples. The most abundant sequence of each OTU was chosen as the representative sequence and classified using the RDP classifier.

Statistical analyses

Of our initial set of 144 samples, three were excluded as outliers at two or more of four measures of OTU alpha diversity (these outliers had alpha diversity values more than three times the interquartile range below the lower 25% percentile). Three additional samples were removed as outliers during initial beta diversity analyses. During rarefaction, 31 samples and 84 samples were removed due to insufficient reads, leaving 107 samples (Table 1) and 54 samples (Table

S1) for 1,500- and 3,000-read datasets respectively. 1,500-read and 3,000-read datasets were compared using the Pearson correlation coefficient and Mantel tests implemented in QIIME pipeline. The two datasets were found to be highly similar (see Results and Discussion); hence in the main text we present the results of the 1,500-read dataset.

Comparison to other mammals. To understand how gut microbiota from the Amboseli baboons compared to other primates and mammals, we conducted PCoA on unweighted Unifrac matrix to compare one randomly selected sample from each adult baboon (n = 14) to 89 individual mammals of 57 species surveyed by (Ley et al., 2008). Only the V1-V3 regions of the 16S rRNA gene were compared. Samples were rarified to 140 reads due to the small number of reads in the mammal dataset (Ley et al., 2008).

Testing predictors of gut microbial alpha diversity. To test which factors best predicted microbial alpha diversity, we constructed linear mixed models of four measures of OTU alpha diversity: OTU richness (i.e. the number of distinct OTUs in a sample), Shannon's H, chao1 (log transformed), and Faith's phylogenetic diversity. All models included host identity as a random factor; the best-fitting models were identified using the log likelihood criterion.

Testing predictors of gut microbial beta diversity. To investigate predictors of gut microbial composition, we first performed exploratory PCoA, followed by hypothesis testing via Canonical Correspondence Analysis (Palmer, 1993). PCoA was performed on unweighted and weighted UniFrac dissimilarities calculated from the relative abundance of OTUs in each sample (Lozupone and Knight, 2005). CCA was performed on the relative abundance of taxa at the phylum, genus, and OTU level and host associated metadata using the vegan package in R (Table 2 and S4). For each test, the best model was selected using the log likelihood criterion, and the significance of each predictor was assessed by permutation tests. We did not correct for multiple comparisons in our CCAs because of the nested nature of these analyses. It would be

overly conservative to account for multiple comparisons because tests of many of the factors (e.g., age, rainfall) are not independent across models. To test which factors predicted the relative abundance of the four most common bacteria phyla, we constructed generalized linear mixed models with host identity as a random factor, and a Poisson-distributed error structure. The best-fitting models were chosen using the log likelihood criterion.

Core microbiota. Since closely related bacterial taxa are sometimes ecologically interchangeable (Harvey and Pagel, 1991), it may be useful to consider phylogenetic relationships when identifying core taxa. Core OTUs were identified using a tree-based algorithm and were defined as those OTUs that belonged to the same lineage and occurred in more than 90% of samples. We identified core OTUs in both baboon and other mammalian (Ley et al., 2008) gut microbiota.

Enterotype analyses. We performed enterotype analysis of the adult baboon gut microbiota as described in (Arumugam et al., 2011), which use Calinski-Harabasz index as an indicator of optimal clustering. In addition, we calculated silhouette coefficient and prediction strength using methods suggested by (Koren et al., 2013) in R (packages: cluster, clusterSim, fpc). Genera that mainly contribute to each enterotype were identified with Randomforest implemented in QIIME. We consider a genus as a main contributor if its removal increase >15% overall estimated generalization error (an estimate of how much error the classifier would have on a novel dataset). Associations between enterotype and age, rainfall, or diet PC axis were tested with Wilcoxon rank sum test. Association between enterotype and host identity, season, sex or social group were tested with Fisher's Exact Test.

Author contributions

MW and EAA conceived the project and designed the experiment. LEG and EAA collected the samples. TR performed DNA extraction, PCR and library preparation. TR and LEG performed data analyses. TR and LEG drafted the manuscript. MW and EAA revised the manuscript and supervised the work. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Zhang Wang and Adam Labonte for help with 16S PCR amplification and Alex Koeppel for initial 16S rRNA data analysis. We are grateful to Jeanne Altmann, who has been a director of ABRP since 1971. ABRP is supported by the National Science Foundation and the National Institute on Aging. In the last decade we acknowledge support from IOS 1053461, IBN 9985910, IBN 0322613, IBN 0322781, BCS 0323553, BCS 0323596, DEB 0846286, DEB 0846532, IOS 0919200, R01AG034513-01, and P01AG031719. We also thank Duke University, Princeton University, the Princeton Center for the Demography of Aging, and the Max Planck Institute for Demography. We thank the Kenya Wildlife Services, the Institute of Primate Research, the National Museums of Kenya, the National Council for Science and Technology, and members of the Amboseli-Longido pastoralist communities for their assistance in Kenya. A number of people contributed to the long-term data collection over the years. Particular thanks go to ABRP's long-term field team (R.S. Mututua, S. Sayialel, and J.K. Warutere) and to T. Wango for his assistance in Nairobi. Karl Pinc provided expertise in database design and management. We also thank our database technicians, particularly D. Onderdonk, C. Markham, T. Fenn, N. Learn, L. Maryott, P. Onyango and J. Gordon. The authors declare not to have any financial or non-financial competing interests.

References

Ahn, J., Sinha, R., Pei, Z., Dominianni, C., Wu, J., Shi, J. et al. (2013) Human gut microbiome and risk of colorectal cancer. Journal of the National Cancer Institute **105**: 1907-1911.

Alberts, S.C., and Altmann, J. (1995) Balancing costs and opportunities: Dispersal in male baboons. Am Nat **145**: 279-306.

Alberts, S.C., and Altmann, J. (2012) The Amboseli Baboon Research Project: 40 years of continuity and change. In Long-term field studies of primates Kappeler, P., and Watts, D.P. (eds): Springer Verlag, pp. 261-287.

Alberts, S.C., Sapolsky, R.M., and Altmann, J. (1992) Behavioral, endocrine, and immunological correlates of immigration by an aggressive male into a natural primate group. Horm behav **26**: 167-178.

Alberts, S.C., Hollister-Smith, J.A., Mututua, R.S., Sayialel, S.N., Muruthi, P.M., Warutere, J.K., and Altmann, J. (2005) Seasonality and long term change in a savannah environment. In Primate Seasonality: Implications for Human Evolution. Brockman, D.K., and Van Schaik, C.P. (eds): Cambridge University Press.

Altmann, S.A. (1973) The pregnancy sign in savannah baboons. J Zoo Anim Med **4**: 8-12.

Amato, K.R. (2013) Co-evolution in context: The importance of studying gut microbiomes in wild animals. Microbiome Science and Medicine **1**: 10.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R. et al. (2011) Enterotypes of the human gut microbiome. Nature **473**: 174-180.

Beehner, J.C., Nguyen, N., Wango, E.O., Alberts, S.C., and Altmann, J. (2006) The endocrinology of pregnancy and fetal loss in wild baboons. Horm behav **49**: 688-699.

Bengmark, S. (2013) Gut microbiota, immune development and function. Pharmaco Res **69**: 87-113.

Bolnick, D.I., Snowberg, L.K., Hirsch, P.E., Lauber, C.L., Knight, R., Caporaso, J.G., and Svanback, R. (2014) Individuals' diet diversity influences gut microbial diversity in two freshwater fish (threespine stickleback and Eurasian perch). Ecol Lett **17**: 979-987.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J. et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad of Sci U S A **108**: 4516-4522.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. Nat methods **7**: 335-336.

Codron, D., Lee-Thorp, J., Sponheimer, M., Ruiter, D., and Codron, J. (2008) What insights can baboon feeding ecology provide for early Hominin niche differentiation? Int J Primatol **29**: 757-772.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E. et al. (2014) Diet rapidly and reproducibly alters the human gut microbiome. Nature **505**: 559-563.

Degnan, P.H., Pusey, A.E., Lonsdorf, E.V., Goodall, J., Wroblewski, E.E., Wilson, M.L. et al. (2012) Factors associated with the diversification of the gut microbial communities within chimpanzees from Gombe National Park. Proc Natl Acad Sci U S A **109**: 13034-13039.

Delsuc, F., Metcalf, J.L., Wegener Parfrey, L., Song, S.J., Gonzalez, A., and Knight, R. (2014) Convergence of gut microbiomes in myrmecophagous mammals. Mol Ecol **23**: 1301-1317.

DeVore, I., and Washburn, S. (1963) Baboon ecology and human evolution. In African Ecology and Human Evolution. FC, H., and F, B. (eds). Chicago: Aldine.

Dillon, R.J., Vennard, C.T., Buckling, A., and Charnley, A.K. (2005) Diversity of locust gut bacteria protects against pathogen invasion. Ecol Lett **8**: 1291-1298.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**: 2194-2200.

Fierer, N., Bradford, M.A., and Jackson, R.B. (2007) Toward an ecological classification of soil bacteria. Ecology **88**: 1354-1364.

Flores, R., Shi, J., Gail, M.H., Gajer, P., Ravel, J., and Goedert, J.J. (2012) Association of fecal microbial diversity and taxonomy with selected enzymatic functions. PLoS ONE **7**: e39745.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics **28**: 3150-3152.

Fulthorpe, R.R., Roesch, L.F., Riva, A., and Triplett, E.W. (2008) Distantly sampled soils carry few species in common. ISME J **2**: 901-910.

Gesquiere, L.R., Wango, E.O., Alberts, S.C., and Altmann, J. (2007) Mechanisms of sexual selection: sexual swellings and estrogen concentrations as fertility indicators and cues for male consort decisions in wild baboons. Horm and behav **51**: 114-125.

Gesquiere, L.R., Learn, N.H., Simao, M.C.M., Onyango, P.O., Alberts, S.C., and Altmann, J. (2011) Life at the top: Rank and stress in wild male baboons. Science **333**: 357-360.

Greenblum, S., Turnbaugh, P.J., and Borenstein, E. (2011) Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc Natl Acad Sci U S A **109**: 594-599.

Harvey, P.H., and Pagel, M.D. (1991) The comparative method in evolutionary biology. Oxford ; New York: Oxford University Press.

Hausfater, G. (1975) Dominance and Reproduction in Baboons: A Quantitative Analysis. Basel: S Karger.

Hong, P.Y., Wheeler, E., Cann, I.K., and Mackie, R.I. (2011) Phylogenetic analysis of the fecal microbial community in herbivorous land and marine iguanas of the Galapagos Islands using 16S rRNA-based pyrosequencing. ISME J **5**: 1461-1470.

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

Iida, N., Dzutsev, A., Stewart, C.A., Smith, L., Bouladoux, N., Weingarten, R.A. et al. (2013) Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. Science **342**: 967-970.

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C.J., Fagerberg, B. et al. (2013) Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature **498**: 99-103.

Koeppel, A.F., and Wu, M. (2012) Lineage-dependent ecological coherence in bacteria. FEMS Microbiol Ecol **81**: 574-582.

Koeth, R.A., Wang, Z., Levison, B.S., Buffa, J.A., Org, E., Sheehy, B.T. et al. (2013) Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med **19**: 576-585.

Koren, O., Knights, D., Gonzalez, A., Waldron, L., Segata, N., Knight, R. et al. (2013) A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. PLoS Comput Biol **9**: e1002863.

Koren, O., Goodrich, Julia K., Cullender, Tyler C., Spor, A., Laitinen, K., Kling Bäckhed, H. et al. (2012) Host remodeling of the gut microbiome and metabolic changes during pregnancy. Cell **150**: 470-480.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S. et al. (2008) Evolution of mammals and their gut microbes. Science **320**: 1647-1651.

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

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

Markle, J.G.M., Frank, D.N., Mortin-Toth, S., Robertson, C.E., Feazel, L.M., Rolle-Kampczyk, U. et al. (2013) Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. Science **339**: 1084-1088.

Martínez, I., Muller, C.E., and Walter, J. (2013) Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. PLoS ONE **8**: e69621.

McKenna, P., Hoffmann, C., Minkah, N., Aye, P.P., Lackner, A., Liu, Z. et al. (2008) The Macaque Gut Microbiome in Health, Lentiviral Infection, and Chronic Enterocolitis. PLoS Pathog **4**: e20.

Moeller, A.H., Degnan, P.H., Pusey, A.E., Wilson, M.L., Hahn, B.H., and Ochman, H. (2012) Chimpanzees and humans harbour compositionally similar gut enterotypes. Nat Commun **3**: 1179.

Moeller, A.H., Li, Y., Mpoudi Ngole, E., Ahuka-Mundeke, S., Lonsdorf, E.V., Pusey, A.E. et al. (2014) Rapid changes in the gut microbiome during human evolution. Proc Natl Acad Sci U S A.

Morgan, X.C., Segata, N., and Huttenhower, C. (2013) Biodiversity and functional genomics in the human microbiome. TIG **29**: 51-58.

Nguyen, N., Gesquiere, L., Alberts, S.C., Altmann, J. (2012) Sex differences in the mother–neonate relationship in wild baboons: social, experiential and hormonal correlates. Anim Behav **83**: 891-903.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. et al. (2013) vegan: Community ecology package. R package version 2.0-7. http://CRAN.R-project.org/package=vegan

Onyango, P.O., Gesquiere, L.R., Altmann, J., and Alberts, S.C. (2013) Puberty and dispersal in a wild primate population. Horm and behav **64**: 240-249.

Palmer, M.W. (1993) Putting things in even better order: the advantages of canonical correspondence analysis. Ecology **74**: 2215-2230.

Philippot, L., Andersson, S.G., Battin, T.J., Prosser, J.I., Schimel, J.P., Whitman, W.B., and Hallin, S. (2010) The ecological coherence of high bacterial taxonomic ranks. Nat Rev Microbiol **8**: 523-529.

Pointing, S.B., Chan, Y., Lacap, D.C., Lau, M.C., Jurgens, J.A., and Farrell, R.L. (2009) Highly specialized microbial diversity in hyper-arid polar desert. Proc Natl Acad Sci U S A **106**: 19964-19969.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids res **41**: D590-596.

Ren, T., Glatt, D.U., Nguyen, T.N., Allen, E.K., Early, S.V., Sale, M. et al. (2013) 16S rRNA survey revealed complex bacterial communities and evidence of bacterial interference on human adenoids. Environ Microbiol **15**: 535-547.

Sapolsky, R.M., and Altmann, J. (1991) Incidence of hypercortisolism and dexamethasone resistance increases with age among wild baboons. Biol Psychiatry **30**: 1008-1016.

Shaikh, A., Shaikh, S., Celaya, C., and Gomez, I. (1982) Temporal relationship of hormonal peaks to ovulation and sex skin deturgescence in the baboon. Primates **23**: 444-452.

Sponheimer, M., Alemseged, Z., Cerling, T.E., Grine, F.E., Kimbel, W.H., Leakey, M.G. et al. (2013) Isotopic evidence of early hominin diets. Proc Natl Acad Sci U S A **110**: 10513-10518.

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.P. et al. (2009) Towards the human intestinal microbiota phylogenetic core. Environ Microbiol **11**: 2574-2584.

Tung, J., Barriero, L.B., Burns, M.B., Grenier, J.C., Lynch, J., Grieneisen, L.E. et al. (in press) Social networks predict gut microbiome composition in wild baboons. eLife.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. Nature **444**: 1027-1131.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E. et al. (2009) A core gut microbiome in obese and lean twins. Nature **457**: 480-484.

Turroni, F., Peano, C., Pass, D.A., Foroni, E., Severgnini, M., Claesson, M.J. et al. (2012) Diversity of bifidobacteria within the infant gut microbiota. PLoS ONE **7**: e36957.

Viaud, S., Saccheri, F., Mignot, G., Yamazaki, T., Daillère, R., Hannani, D. et al. (2013) The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. Science **342**: 971-976.

von Mering, C., Hugenholtz, P., Raes, J., Tringe, S.G., Doerks, T., Jensen, L.J. et al. (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. Science **315**: 1126-1130.

Wang, J., Linnenbrink, M., Kunzel, S., Fernandes, R., Nadeau, M.J., Rosenstiel, P., and Baines, J.F. (2014) 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 **111**: E2703-2710.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol **73**: 5261-5267.

Wildt, D.E., Doyle, L.L., Stone, S.C., and Harrison, R. (1977) Correlation of perineal swelling with serum ovarian hormone levels, vaginal cytology, and ovarian follicular development during the baboon reproductive cycle. Primates **18**: 261-270.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A. et al. (2011) Linking long-term dietary patterns with gut microbial enterotypes. Science **334**: 105-108.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M. et al. (2012) Human gut microbiome viewed across age and geography. Nature **486**: 222-227.

Yildirim, S., Yeoman, C.J., Sipos, M., Torralba, M., Wilson, B.A., Goldberg, T.L. et al. (2010) Characterization of the fecal microbiome from non-human wild primates reveals species specific microbial communities. PLoS ONE **5**: e13963.

Tables and figures

Table 1. Sample size information, including the number of individuals and fecal samples used in analyses of the dataset rarefied to 1,500 reads.

Individual	Sex	Number of fecal samples	Range of years samples were collected	Age range or age at time of sample collection (years)
BEAM	M	10	1994 - 2001	5.95 - 13.16
DUNLIN	F	10	1996 - 1999	0.72 - 3.69
OKOT	M	10	1996 - 1998	1.32 - 2.81
LEBANON	M	8	1997 - 2009	0.17 - 12.16
OCEAN	M	8	1997 - 2000	0.6 - 3.78
VIXEN	F	8	1994 - 1998	17.06 - 20.84
DRONGO	F	7	1996 - 2009	6.99 - 19.88
ECHO	F	7	1995 - 2001	3.43 - 9.39
VANGA	M	7	1995 - 2001	2.54 - 9.31
GOLON	M	6	1996 - 1999	17.76 - 20.63
LAWYER	M	6	2001 - 2001	1.12 - 1.98
OXYGEN	F	6	2000 - 2001	1.05 - 2.43
HONEY	F	3	1999 - 2000	1.85 - 3.13
AMIGO	M	1	1998	15.07
CABANA	F	1	1999	0.53
CEDAR	M	1	1998	2.86
DYNAMO	M	1	1998	0.94
HEKO	F	1	1997	14.27
LADHA	F	1	1998	4.57
LARK	F	1	1997	9.71
LAZA	F	1	1998	6.9
PLATO	M	1	1998	8.6
VOGUE	F	1	1998	1.08
VORTEX	F	1	1997	10.32

Table 2. CCA analysis of environment and host traits that predicted variation in gut microbial community composition rarefied to a level of 1,500 reads.

Dataset	Num of samples	Factors included in the model	Significant factors at		
			phylum level	genus level	OTU level
Main dataset	107	age, rainfall, sex, individual ID, social group, natal social group, group size	rainfall (P=0.02), age (P=0.01)	rainfall (P=0.02), age (P=0.02)	None
Subset with diet information	76	age, rainfall, sex, diet diverisity (richness, Shannon's H or PCoA axis), individual ID	rainfall (P=0.02), age (P=0.04), diet Shannon's H (P=0.05) or diet PC1 (P=0.02)	rainfall (P=0.03), age (P=0.05), diet Shannon's H (P=0.02) or diet PC1 (P=0.02)	None
Infant	22	age, rainfall, sex, individual ID	sex (P=0.04)	sex (P=0.03)	None
Juvenile	38	age, rainfall, sex, individual ID	None	None	None
Infant / juvenile	60	age, rainfall, sex, suckle status, individual ID	None	None	None
Adult male with rank information	21	age, rainfall, adult male rank, natal social group, individual ID	rainfall (P=0.05)	rainfall (P=0.03)	None
Adult female with rank information	24	age, rainfall, adult female rank, reproductive status, individual ID	rainfall (P=0.14)	rainfall (P=0.02)	None

Table 3. Best-supported generalized linear mixed models (Poisson-link) explaining variation in abundance of the four most common bacteria phyla for the main dataset (n = 107 samples). Host identity is modeled as a random effect.

Bacteria phylum	**Fixed effects**	**Estimate**	**SE**	**Z**	***P* -value**
Actinobacteria	age	-0.024	0.004	-6.599	<0.001*
	rainfall	-0.003	0.0001	-17.897	<0.001*
Bacteroidetes	age	-0.027	0.005	-4.918	<0.001*
Firmicutes	age	0.008	0.002	4.44	<0.001*
	rainfall	0.003	<0.0001	37.21	<0.001*
Proteobacteria	age	-0.0299	0.008	-3.69	<0.001*

Figure 1. Phylum level bacterial composition across 107 samples from 24 individual baboons. Each column represents one fecal sample. Y-axis values represent the relative abundance of each phylum classified by RDP classifier. Samples are sorted by the relative abundance of Actinobacteria in the sample.

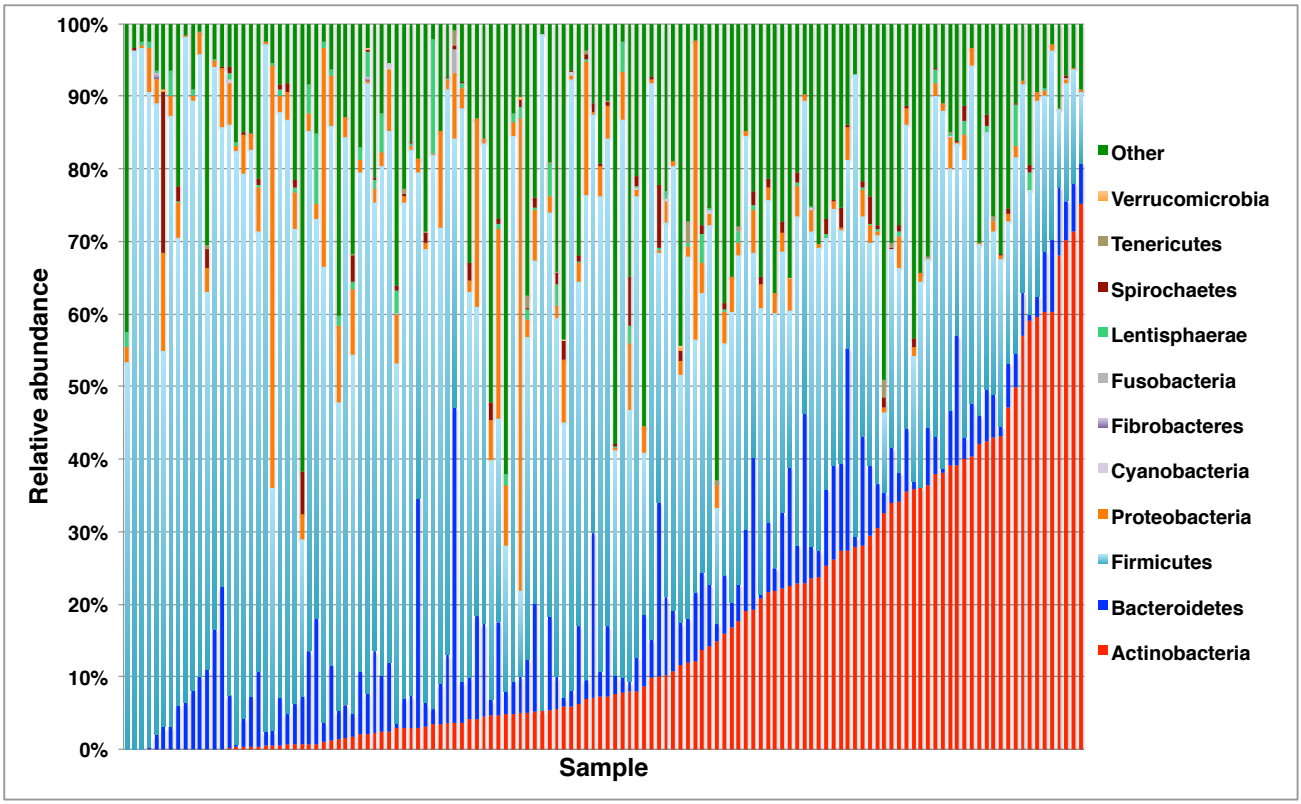
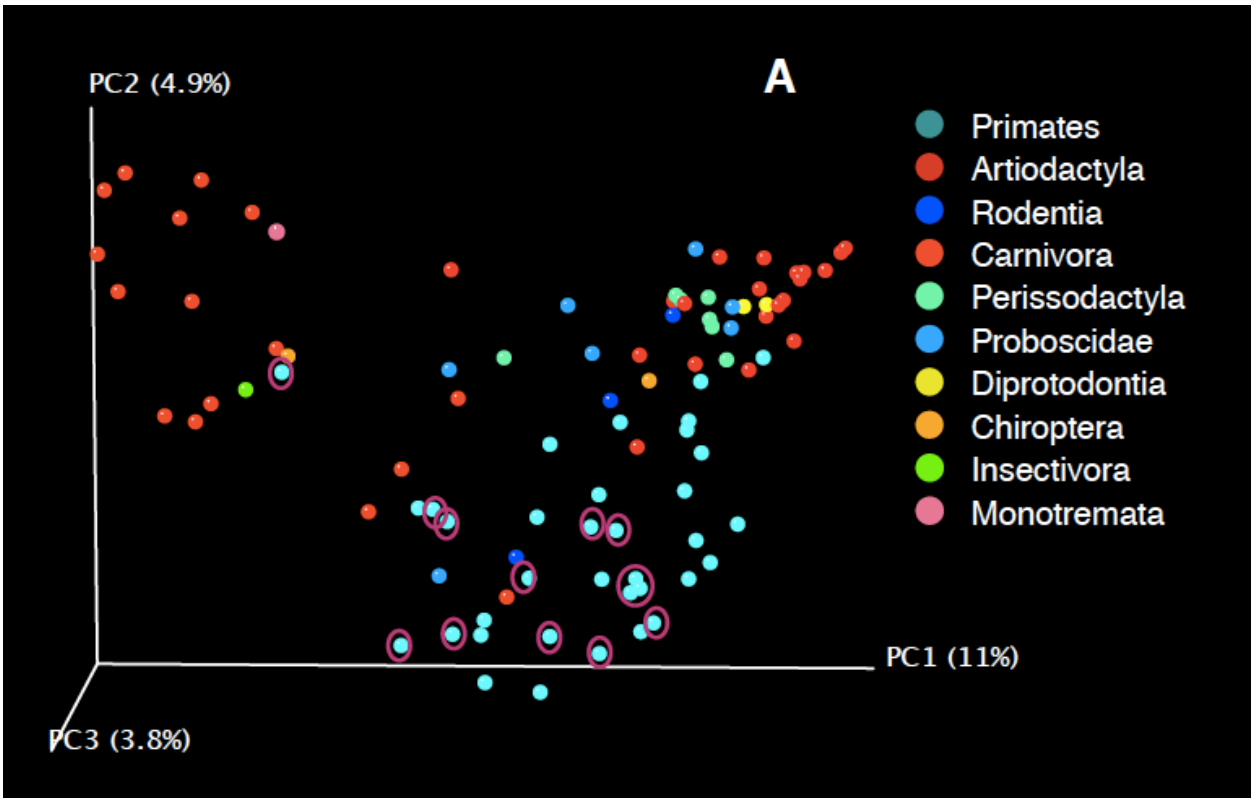


Figure 2. PCoA analysis of the weighted UniFrac dissimilarities comparing gut microbiota of baboons to other mammals. Each point corresponds to a sample colored by (A) host taxonomy and (B) host diet type. Baboon samples are circled in red; the 14 baboons samples were drawn at random representing one each from the 14 adult individuals (AMIGO, BEAM, DRONGO, ECHO, GOLON, HEKO, LADHA, LAZA, LARK, LEBANON, PLATO, VANGA, VIXEN, VORTEX) included in our dataset. The percentage of the variation explained by the plotted principal coordinates is indicated on the axes.



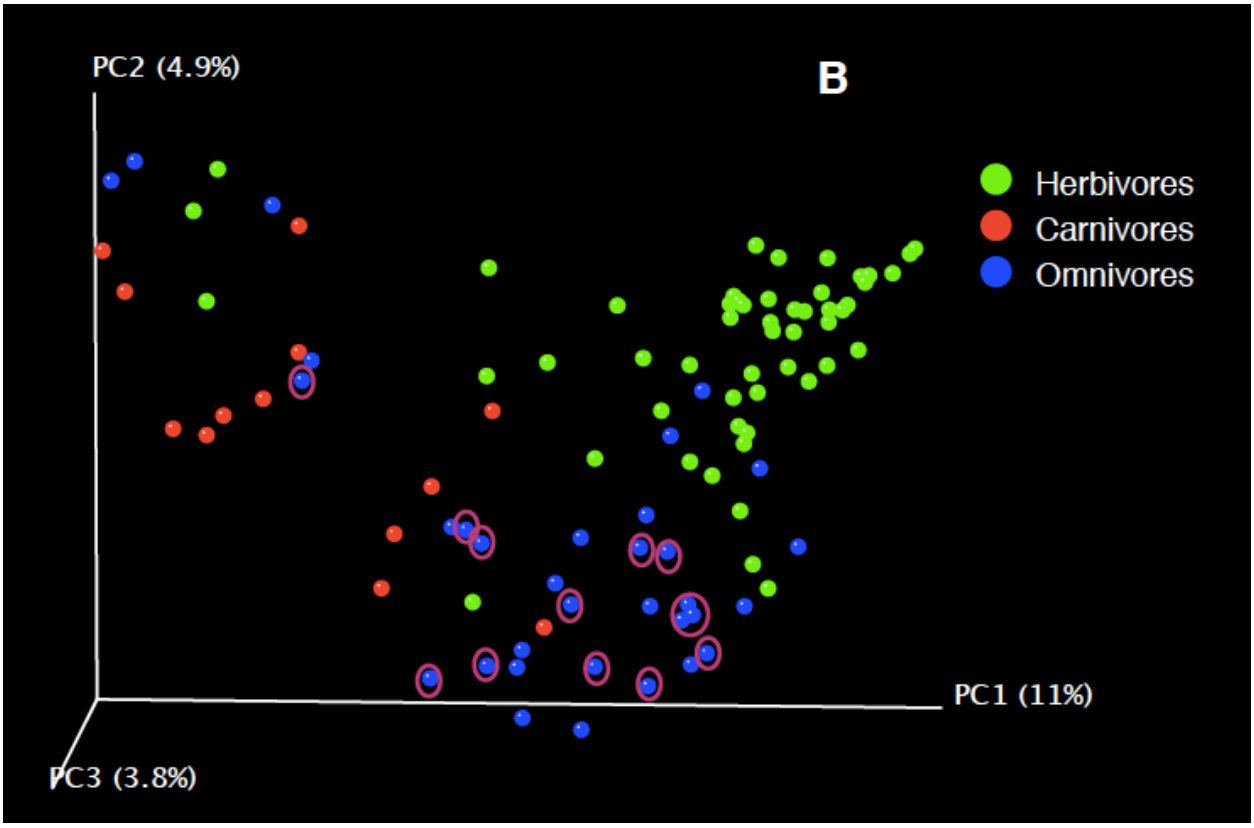

PC2 (4.9%)
B
Herbivores
Carnivores
Omnivores
PC1 (11%)
PC3 (3.8%)

Figure 3. The relationship between baboon age (in years) and gut microbiota alpha diversity as measured by OTU Shannon's H. Infant and juvenile baboons had higher variance in Shannon's H than adults (Brown-Forsyth test $F_{(1, 104.876)} = 6.988$, $P = 0.009$)

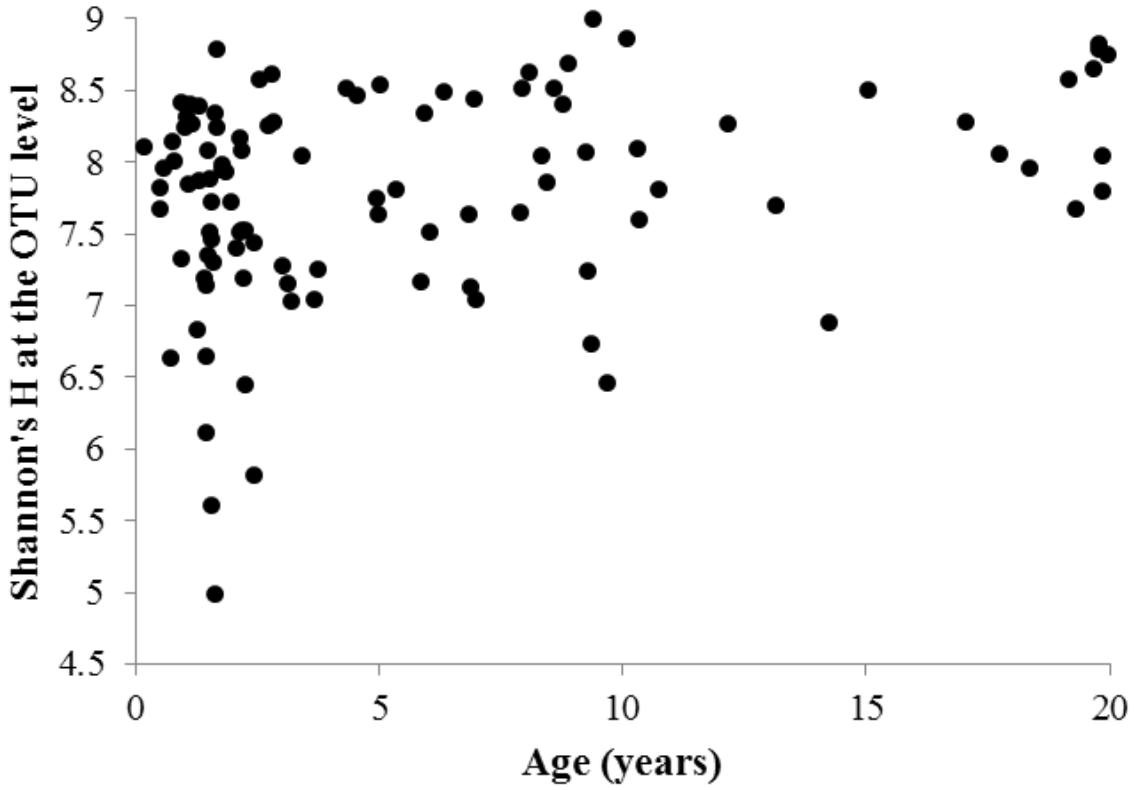


Figure 4. A time-decay plot of gut microbiota dissimilarity. Each dot represents a comparison between two samples of the same baboon collected at different time points, with different marker colors representing different baboons. X-axis represents the time span (in days) between the sample collection times. Y-axis represents the weighted UniFrac dissimilarity. (A) 5000 days. (B) 365 days. Correlation between sampling time span and microbiota weighted Unifrac dissimilarity for each individual is summarized in Table S6.

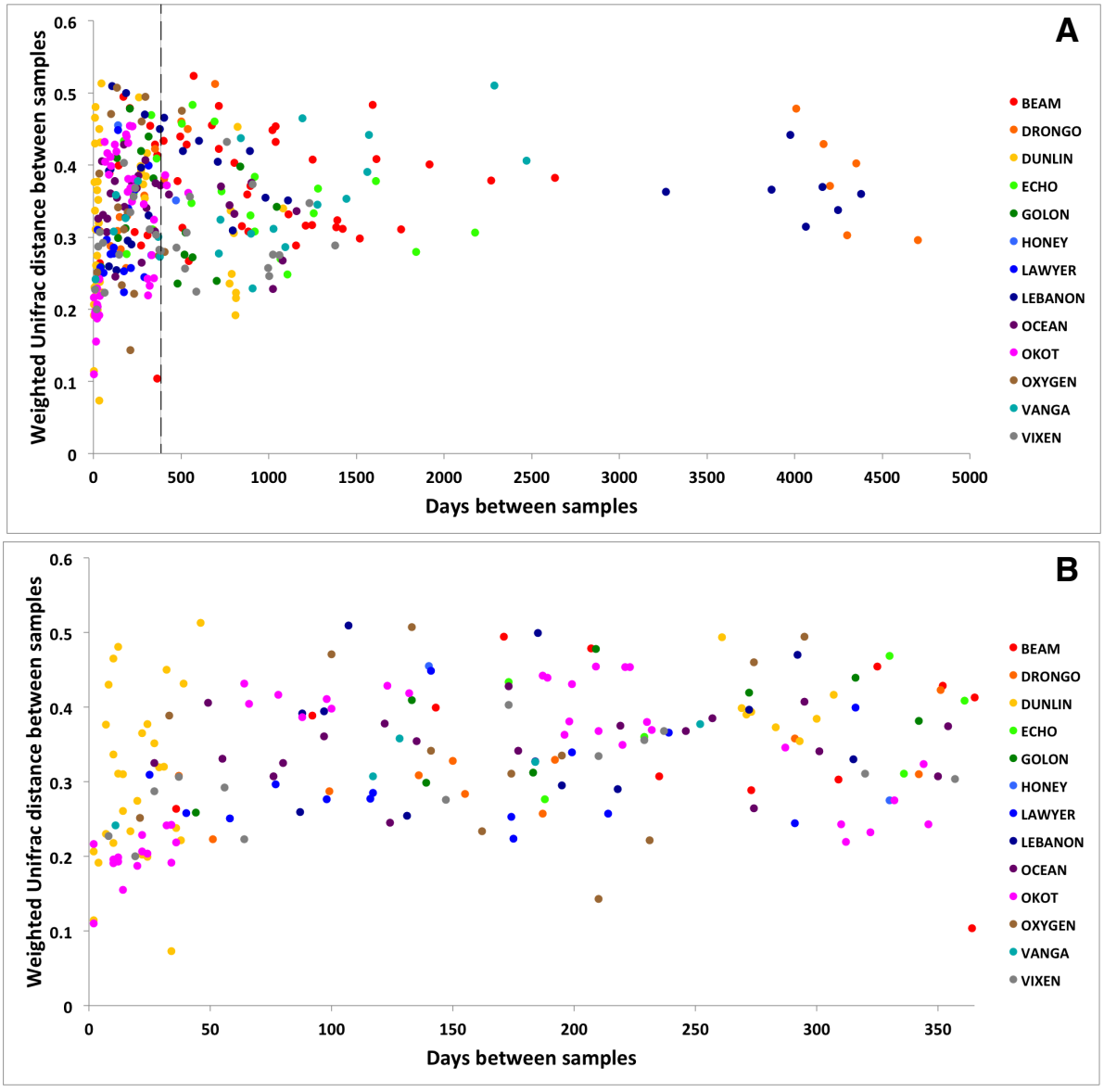


Figure 5. PCoA visualization of baboon enterotypes (ellipses) identified by PAM clustering. Black dots represent abundance distributions of bacterial genera from an individual host and numbered white rectangles mark the center of each enterotype. Bacterial genera that mainly contribute to each enterotype are listed.

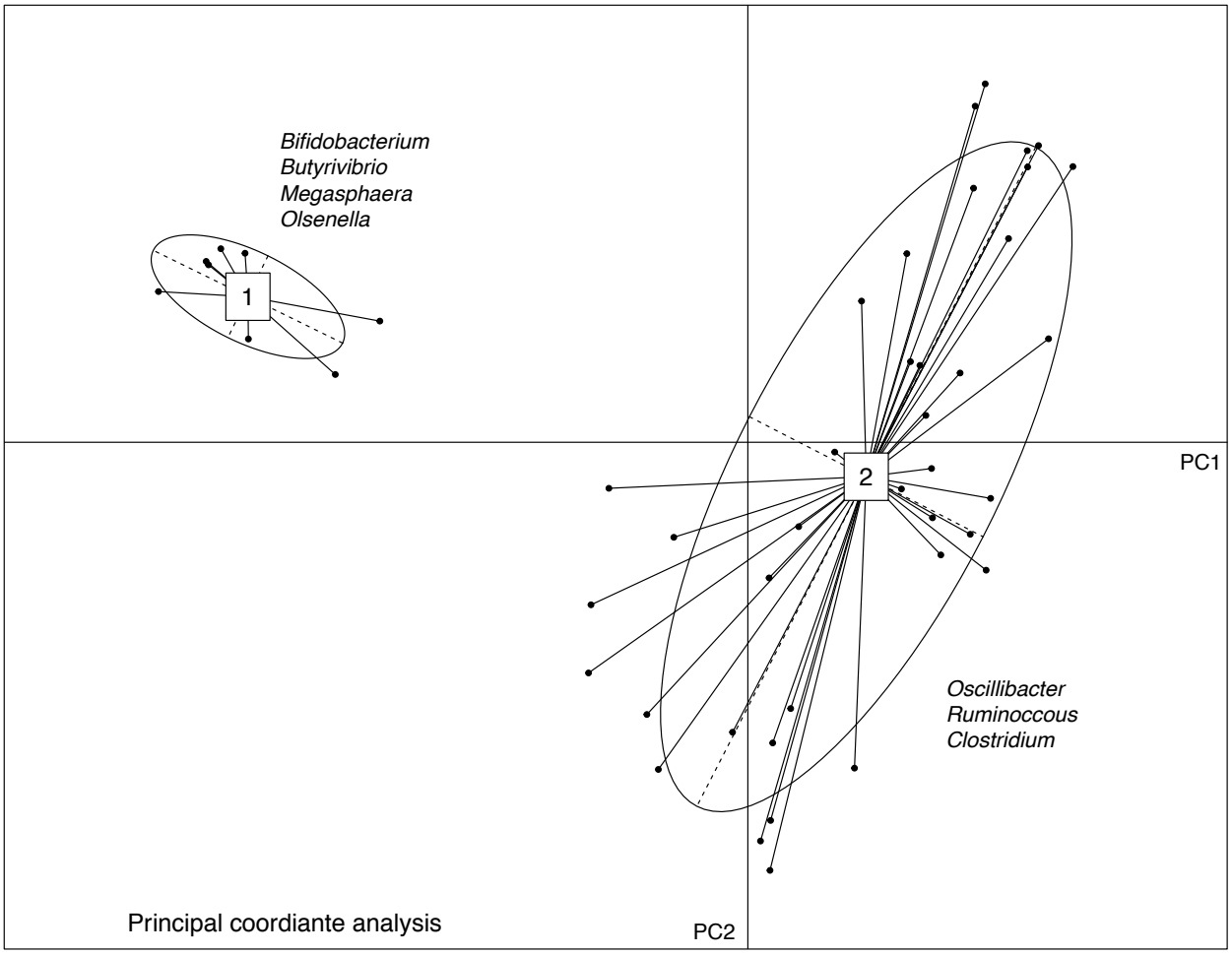
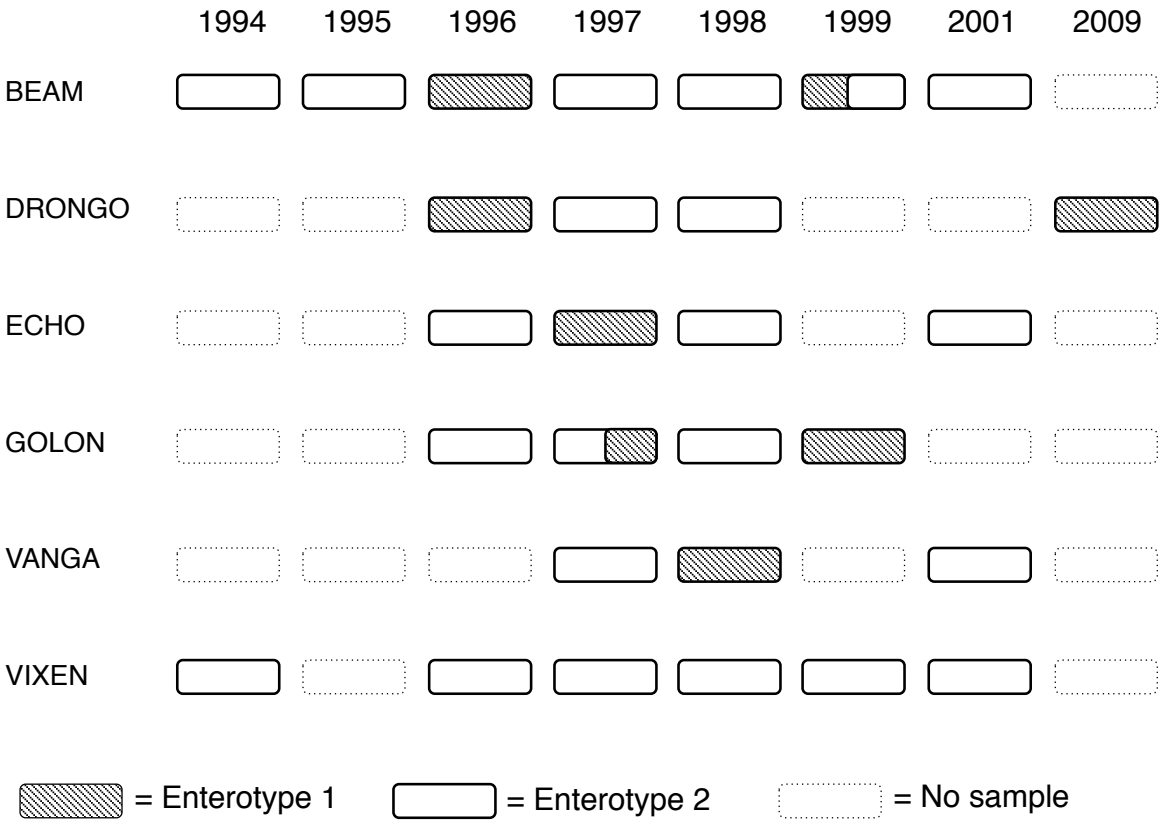


Figure 6. Baboon enterotypes switched over time. Samples were collected from 1994 to 1999, and in 2001 and 2009. 8 (of 14) adults only had one sample and therefore were not shown here. Filled rectangles: enterotype 1; unfilled rectangles: enterotype 2; rectangles with dashed line: sample missing. When an individual switched enterotypes during the middle of the year, it is represented by a hybrid rectangle (half filled and half unfilled).



Supplementary information

Supplementary tables

Table S1. Sample size information, including the number of individuals and fecal samples used in analyses of the dataset rarefied to 3,000 reads.

Individual	Sex	Number of samples	Range of years samples were collected	Age range or age at time of sample collection (years)
BEAM	M	9	1994 - 2001	5.95 - 13.16
DUNLIN	F	8	1996 - 1997	0.72 - 1.56
OCEAN	M	5	1997 - 2000	0.6 - 3.78
OKOT	M	5	1996 - 1998	1.32 - 2.81
VANGA	M	5	1995 - 1998	3.05 - 6.05
DRONGO	F	3	1996 - 1997	6.99 - 8.09
GOLON	M	3	1997 - 1999	19.19 - 20.63
LAWYER	M	3	2001 - 2001	1.19 - 1.98
HONEY	F	2	1999 - 2000	1.85 - 2.75
LEBANON	M	2	1998 - 2000	1.57 - 3.21
OXYGEN	F	2	2001 - 2001	1.62 - 2.16
VIXEN	F	2	1994 - 1997	17.06 - 19.97
DYNAMO	M	1	1998	0.94
ECHO	F	1	1997	5.88
HEKO	F	1	1997	14.27
LARK	F	1	1997	9.71
VOGUE	F	1	1998	1.08

Table S2. Unweighted UniFrac dissimilarity comparison within and between mammalian orders or diet types.

Group	Average within group dissimilarity	Average between group dissimilarity	*P* value (Wilcoxon rank sum test)
Order	0.80	0.86	2E-16
Diet type	0.82	0.87	2E-16

Table S3. Diet items included in each diet category.

Diet category	Diet item
Grass	Grass corms (all species)
	Grass leaves (all species)
	Grass blade bases (all species)
	Grass seed head (all species)
Gum	Gum from *Acacia xanthophloea*
Leaves	*Lyceum sp.* leaves
	Azima tetracantha leaves
	Acacia xanthophloea leaves
	Salvadora persica leaves
	Suaeda monoica leaves
	Tribulus terrestris leaves
Fruits	*Trianthema ceratosepala* fruits
	Azima tetracantha fruits
	Abutelon sp. fruits
	Lyceum sp. fruits
	Ramphicarpa montana fruits
	Salvadora persica fruits
	Solanum dubium fruits
	Tribulus terrestris fruits
	Withania sp. fruits
Blossoms	*Acacia xanthophloea* blossoms
	Ramphicarpa montana blossoms
	Acacia tortilis blossoms
Bark	Bark from *Acacia xanthophloea*
Pods	Fresh, green seed pods of *Acacia* spp.
Seeds	Dried seeds of *Acacia* spp.
Invertebrates	Invertebrates of unknown species
Dung	Liquid from or items in elephant dung
Unknown	Unknown diet items (i.e. those that could not be seen by observers)

Table S4. CCA analysis of environment and host factors for the 3,000 read dataset.

Dataset	Number of samples	Factors tested	Best model at		
			phylum level	genus level	OTU level
Full dataset	54	age, rainfall, sex, individual ID, social group, natal social group, group size	rainfall (P=0.12), age (P=0.13)	rainfall (P=0.09)	None
Subset with diet diversity info	38	age, rainfall, sex, diet diverisity (richness, Shannon's H or PCoA axis), individual ID	None	rainfall (P=0.08), diet PC1 (P=0.05)	None

Table S5. Best-supported generalized linear mixed model (Poisson-link) explaining variation in abundance of the four most common bacteria phyla for the subset of 76 samples with diet data. Individual identity is a random effect in all models.

Bacteria phylum	Fixed effect	estimate	S.E.	Z	p-value
Actinobacteria	rainfall	-0.005	0.000	-15.126	<0.001*
	diet PC1	0.248	0.041	6.086	<0.001*
	diet PC2	1.858	0.113	16.403	<0.001*
	diet PC3	1.981	0.097	20.365	<0.001*
Bacteroidetes	rainfall	-0.002	0.000	-5.101	<0.001*
	diet PC1	-0.849	0.073	-11.70	<0.001*
	diet PC2	-0.537	0.106	-5.048	<0.001*
	diet PC3	1.044	0.161	6.471	<0.001*
Firmicutes	age	-0.006	0.002	-2.68	0.007*
	rainfall	0.001	0.000	13.53	<0.001*
	diet PC1	-0.578	0.026	-22.60	<0.001*
	diet PC2	-0.770	0.034	-22.57	<0.001*
	diet PC3	-0.549	0.056	-9.78	<0.001*
Proteobacteria	age	0.041	0.011	3.91	<0.001*
	diet PC1	3.270	0.071	46.00	<0.001*
	diet PC2	-0.706	0.089	-7.96	<0.001*

Table S6. Mantel test of correlation between sampling time interval and microbiota weighted Unifrac dissimilarity between samples that were collected from the same individual.

Individual	Number of samples	Mantel *r* statistic	*P* value
BEAM	10	-0.04	0.882
DUNLIN	10	-0.03	0.899
OKOT	10	0.40	0.095
OCEAN	8	-0.27	0.263
VIXEN	8	0.11	0.696
DRONGO	7	0.23	0.452
ECHO	7	-0.41	0.180
VANGA	7	0.61	0.038*
GOLON	6	-0.17	0.692
LAWYER	6	0.25	0.365
OXYGEN	6	0.10	0.792
HONEY	3	-0.64	0.481

Supplementary Figures

Figure S1. The distribution patterns of the unclassified and classified OTU at the phylum level across the samples. OTUs have been sorted into bins based on their prevalence in the samples (X-axis). Y-axis is the count of OTUs in each bin.

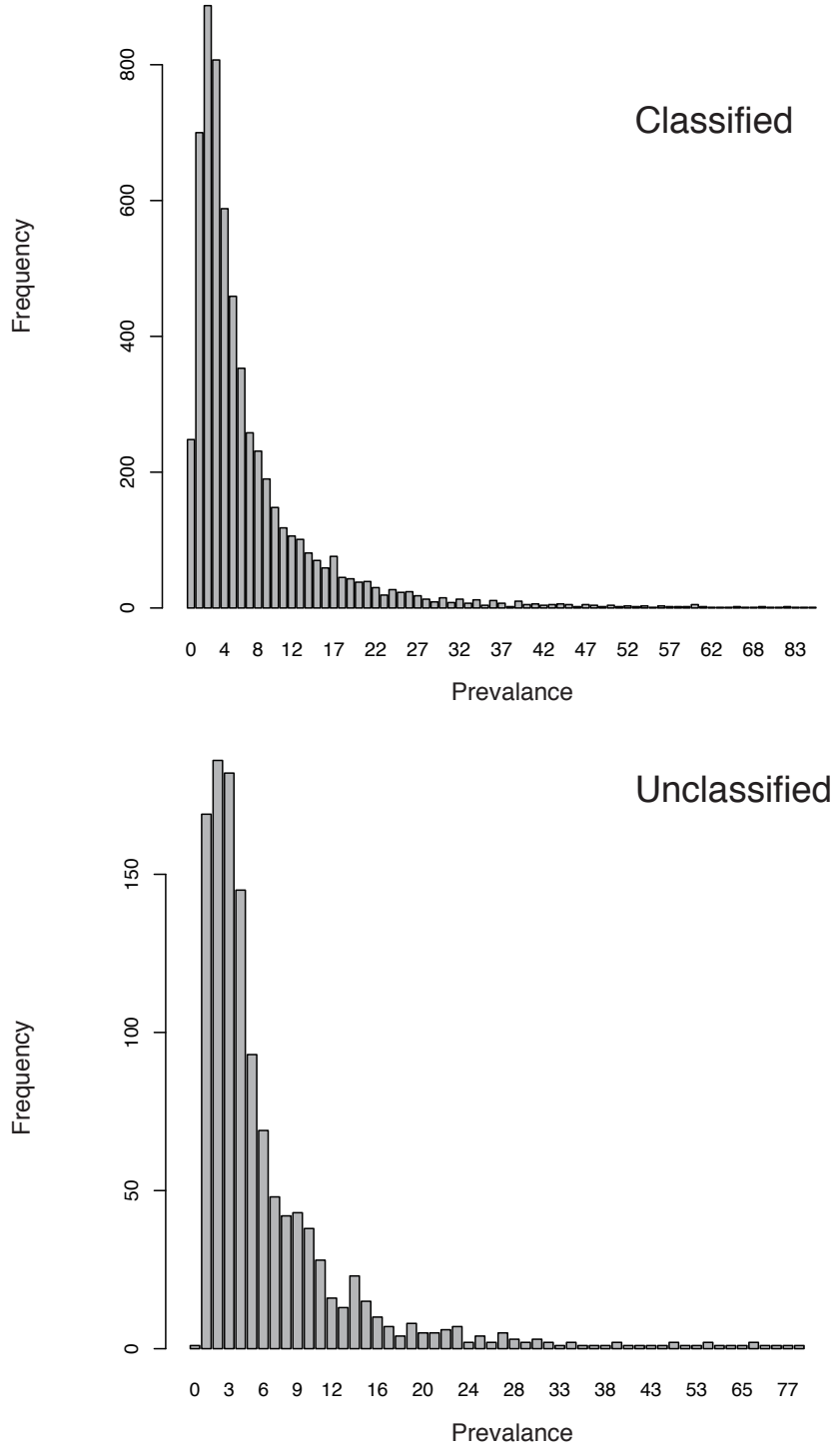


Figure S2. PCoA analysis of the weighted UniFrac dissimilarities comparing baboon gut microbiota. Each point corresponds to a sample colored by (A) individual identity, (B) sex, (C) ageclass and (D) season, (E) diet group. Baboons with diet composition information (n = 76) were divided into 3 diet groups by the relative abundance of grass, fruit and invertebrate in their diet guided by the PCoA plot of diet Bray-Curtis dissimilarity: 1. Fruit, if fruit percentage is >=20%; 2. Invertebrate, if there is invertebrate in diet; 3. Grass, if grass percentage is >=70%.

Figure S2 is available at:

http://onlinelibrary.wiley.com/wol1/doi/10.1111/1462-2920.12852/suppinfo

Figure S3. The first principal coordinate of variation in diet composition (diet PC1) as a function of the 11 primary diet components (Table S3). Blue lines represent lowess regression fits. PC1 explained 46% of the variation in diet composition and is associated with a tradeoff in the proportion of grass (-) versus fruit (+) in the baboons' diets.

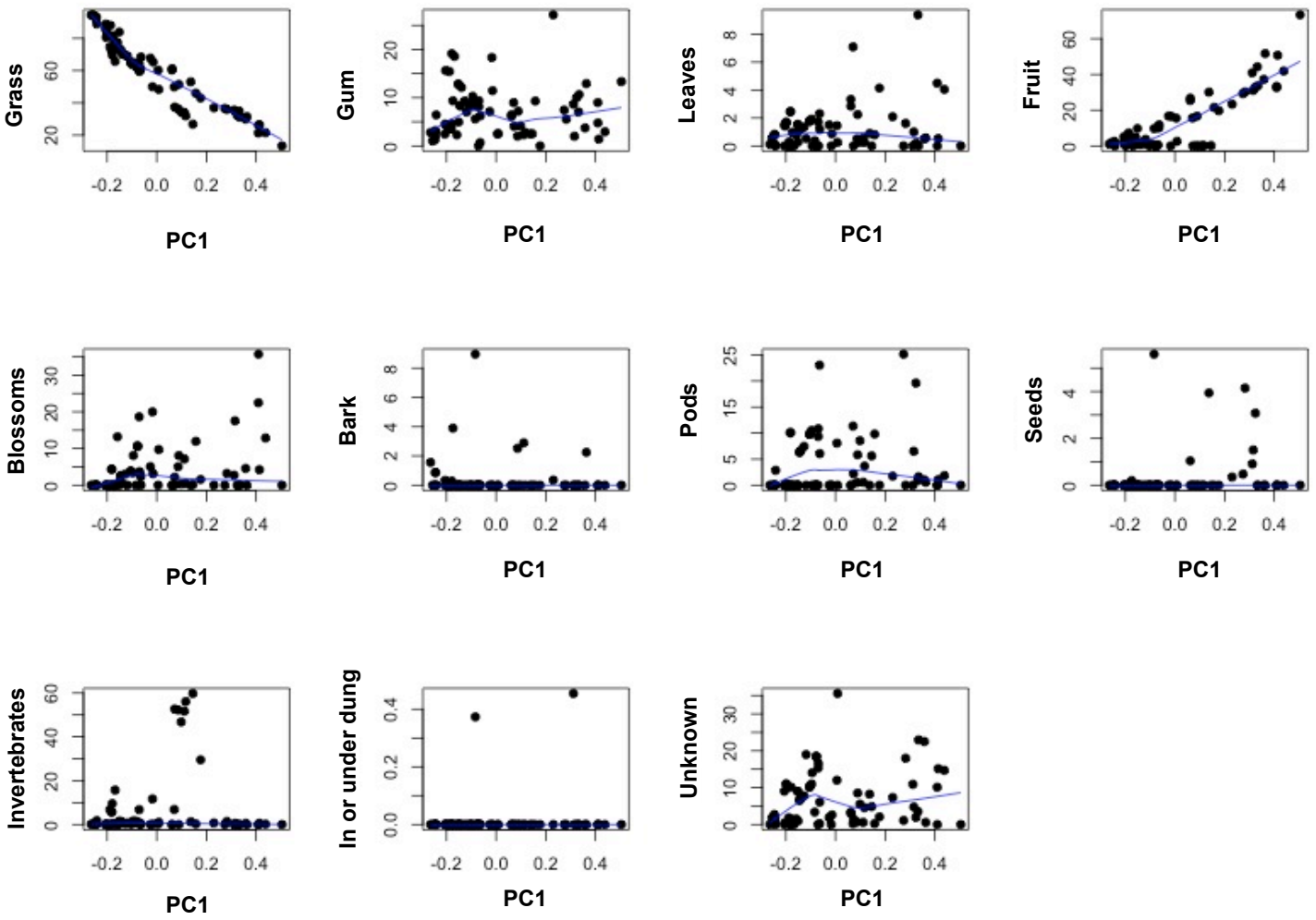


Figure S4. The second principal coordinate of variation in diet composition (diet PC2) as a function of the 11 primary diet components (Table S3). Blue lines represent lowess regression fits. PC1 explained 23% of the variation in diet composition and is associated with a tradeoff in proportion of insects (-) versus fruit (+) in the baboons' diets.

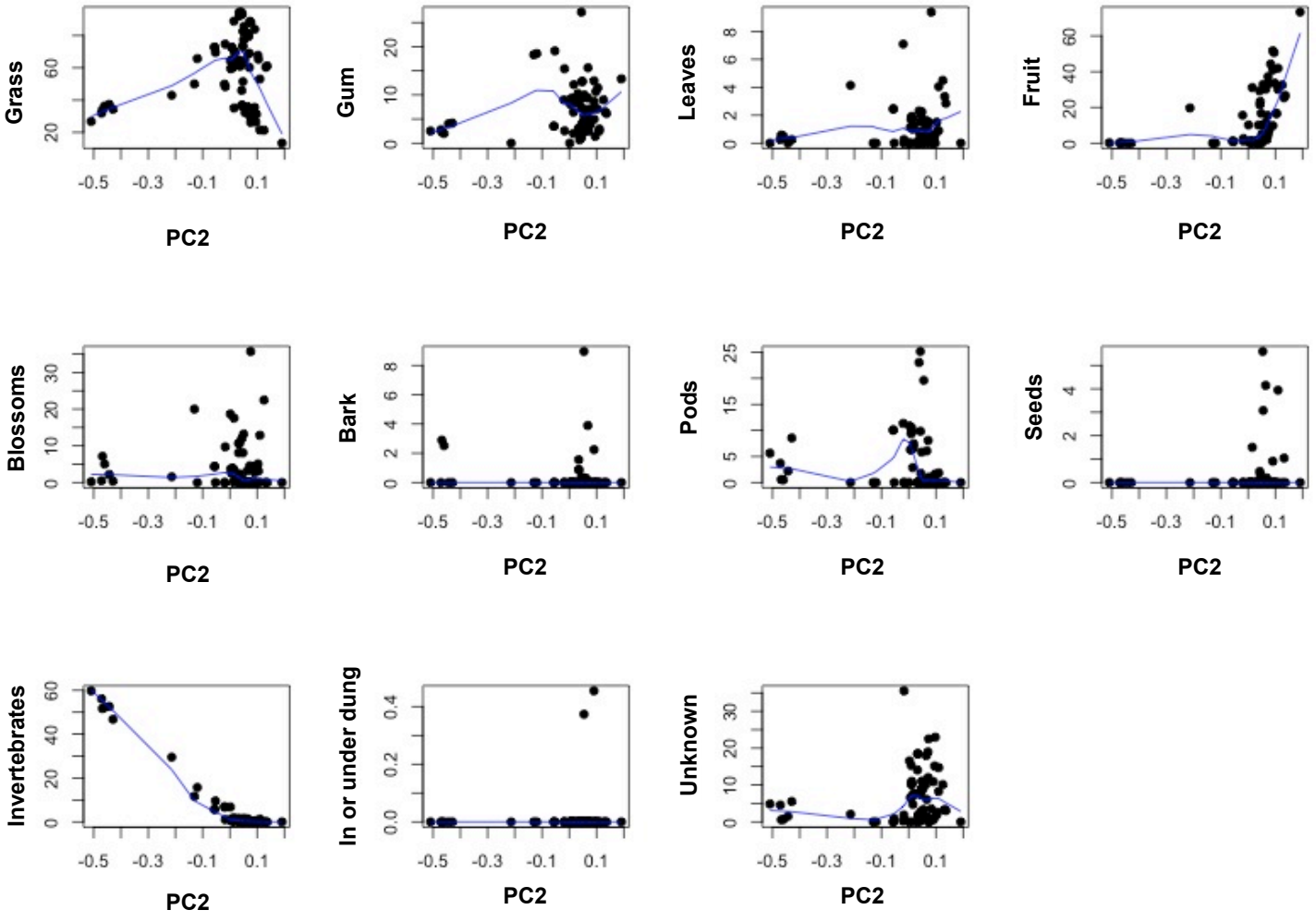
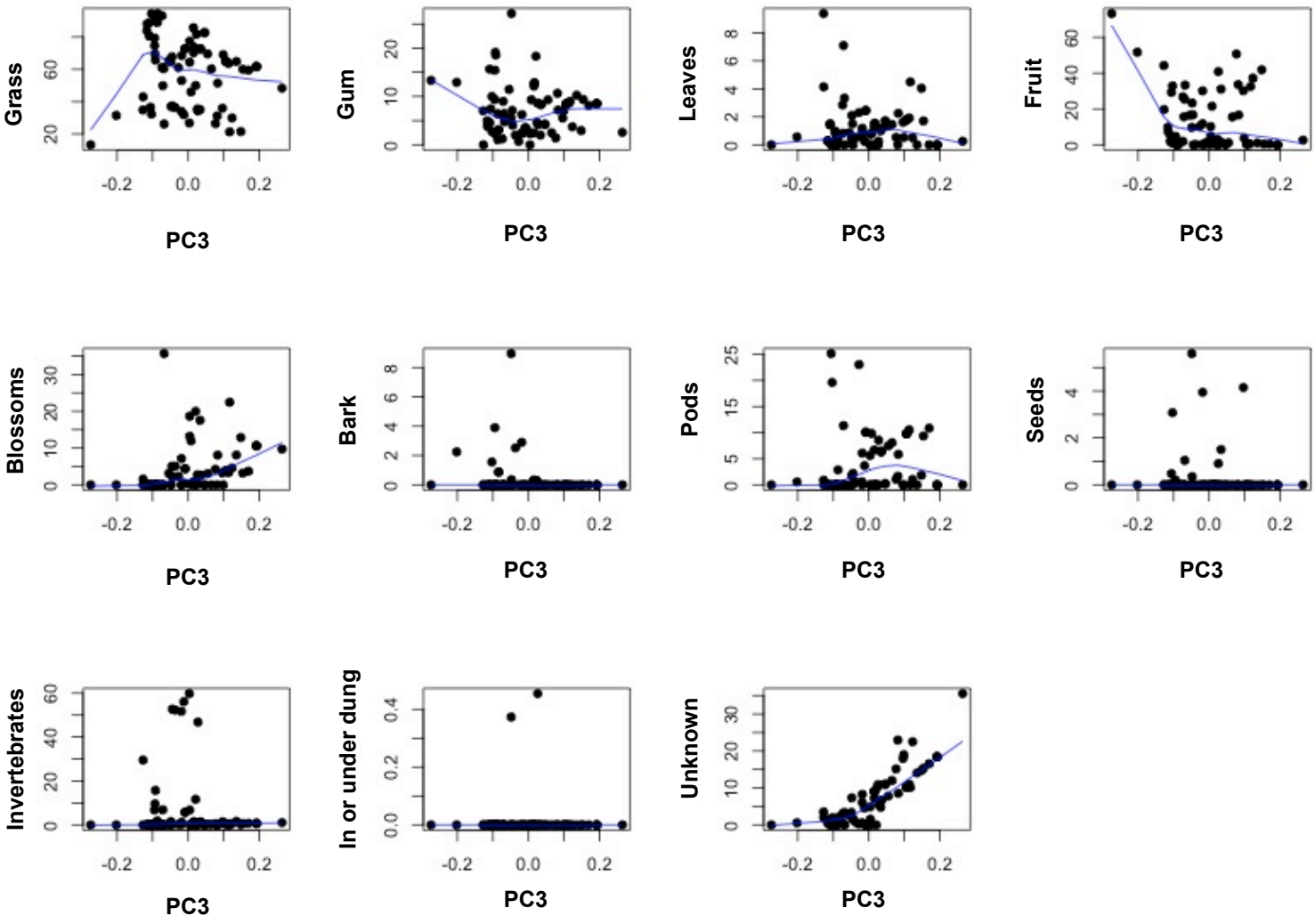


Figure S5. The third principal coordinate of variation in diet composition (diet PC3) as a function of the 11 primary diet components (Table S1). Blue lines represent lowess regression fits. PC1 explained 11% of the variation in diet composition and is associated with the proportion of the diet attributed to 'unknown' categories (-).



Chapter 3. Remarkable seasonal oscillations in red squirrel gut microbiota

Formatted as a co-authored manuscript: Ren T, Dantzer B, McAdam AG and Wu M.

Abstract

Gut micorbiota contribute significantly to host biology and health. At present, our understanding of gut microbiota has been limited primarily to findings from human and laboratory animals, and what ecological factors shape the gut microbiota structure in hosts' natural environment remains largely unexplored. To fill this gap, we conducted a comprehensive study of gut microbiota of a well-studied red squirrel population. Red squirrels are territorial, solitary and lived in a highly seasonal environment, and therefore represent a very attractive system to study the temporal and spatial dynamics of gut microbiota. For the first time, this study revealed significant spatial patterns of gut microbiota within a host population, suggesting limited dispersal could play a role in shaping and maintaining the structure of gut microbial communities. We also found a remarkable seasonal rhythm in red squirrel's gut microbial composition manifested by a tradeoff between relative abundance of genera *Oscillospira* and *Corpococcus*, and clearly associated with seasonal variation in diet availability. Despite strong environmental effects, we found evidences of individuality and maternal effect, but host genetics does not seem to be a significant driver.

Introduction

Mammalian guts harbor trillions of microbes of thousands of species, which play important roles in diverse aspects of host biology, including nutrition, immune system development and behavior. Changes in gut microbial composition have been linked to a plethora of health and diseases (Qin *et al.*, 2012; Morgan *et al.*, 2012; Hooper *et al.*, 2012; Markle *et al.*, 2013). Previous studies have shown that host diet, age, sex, genetics, environmental exposure all drive normal gut microbial variation (Yatsunenko *et al.*, 2012; Claesson *et al.*, 2012; Markle *et al.*, 2013; Goodrich *et al.*, 2014; Lee *et al.*, 2014; Ren *et al.*, 2015; David *et al.*, 2015; Martínez *et al.*, 2015; Amato *et al.*, 2016). However, to date, most studies have been focused on human population and laboratory animals in controlled settings and much remains to be elucidated about the ecological forces shaping gut microbial diversity and their relative strengths in nature. Studies of wild animal populations can provide important insights into how complex environment, host biology and their interactions affect gut microbiota in hosts' natural habitats where hosts and microbes have coevolved.

Diet is believed to be a key selective factor in shaping gut microbiota in wild animals on the evolutionary scale. For example, large differences in gut microbial communities have been found among carnivorous, herbivorous and omnivorous mammals (Ley *et al.*, 2008; Delsuc *et al.*, 2014). Different fruit fly species feeding on distinct diets (fruits vs flowers) differ in their gut microbiota (Chandler *et al.*, 2011). On a shorter ecological timescale, wild animals often face temporal variation in food availability and shift their diet accordingly. However, little is known about how much this variation influences gut microbiota. It has been suggested that seasonal variations in gut microbial composition found in wild mice (Maurice *et al.*, 2015), ground squirrel (Carey *et al.*, 2013) and giant panda (Xue *et al.*, 2015) are largely driven by the seasonal

shifts in diet composition. Nonetheless, these studies either lacked detailed dietary information in the wild or were performed on wild animals in captivity. Further research is necessary to test if and the extent to which seasonal dietary shift shapes gut microbiota in wild animals.

Host genetics can also play a role in controlling gut microbial community structure. Accumulating evidence has linked specific host genetic loci to gut microbial variation in human and mice (Khachatryan *et al.*, 2008; Benson *et al.*, 2010; Rausch *et al.*, 2011; Frank *et al.*, 2011; Rehman *et al.*, 2011; Wacklin *et al.*, 2011; McKnite *et al.*, 2012). Although early studies suggested otherwise (Turnbaugh *et al.*, 2009; Yatsunenko *et al.*, 2012), a recent large-scale study comparing human monozygotic and dizygotic twins revealed significant host genetic effect on gut microbial diversity (Goodrich *et al.*, 2014). In contrast, little is known about the contribution of host genetics in shaping gut microbiota in wild animals, as this type of study depends on knowledge of genetic relatedness within a host population.

While many studies focused on important deterministic factors, little attention has been paid to the role of stochastic processes such as dispersal on structuring gut microbiota. With limited dispersal in homogenous environment, we would predict that hosts living together should exhibit more similar communities of gut microbes than those living further apart. Geographical variations of gut microbiota have been observed in humans. For instance, family members have higher degree of gut microbiota similarity than unrelated individuals, and distinct gut microbial communities were found in populations living on different continents (Turnbaugh *et al.*, 2009; Lee *et al.*, 2011; Yatsunenko *et al.*, 2012; Tims *et al.*, 2013). Likewise, a recent study found biogeographic variation in wild mice populations (Linnenbrink *et al.*, 2013). However, these patterns do not necessarily indicate dispersal limitation because they can be attributed to closer genetic relatedness or shared common environment factors as well.

Furthermore, most studies focused on comparisons between populations separated on large distance scales, but few have investigated the role of dispersal limitation within a host population, where environment is expected to be more homogeneous and less of a concern as a confounding factor.

One particular important mechanism of gut microbiota dispersal is through mother-offspring transmission. Mother can provide the initial inoculum for the gut microbiota in mammalian newborns. For example, mother koalas produce "fecal paps", which contains the bacteria necessary to digest gut leaves, and feed them to the young (Osawa et al., 1993). Accordingly, strong kinship effects were found in several studies (Ley *et al.*, 2005; Lucas and Heeb, 2005; C. Palmer *et al.*, 2007; Perez *et al.*, 2007) where the gut microbiota of offspring were more similar to their mother's than those of unrelated individuals. However, it is not clear how much of the similarity was due to genetics and how much was due to maternal effect.

We performed a large-scale study on a well-characterized wild North American red squirrel (*Tamiasciurus hudsonicus*) population to assess the relative contribution of environmental factors, host genetics, maternal effect and geography to the diversity of gut microbiota. As part of the Kluane Red Squirrel Project, every red squirrel in the population has been continuously monitored year around since 1987 and multiple environmental and host factors were recorded, including age, sex, territory membership, dietary composition and pedigree. Red squirrels live in a strongly seasonal environment and as a result experience recurrent seasonal fluctuations in their diet. Red squirrels defend their exclusive territories over lifetime and thus spend most time in solitude. As such, red squirrels represent a very unique and attractive system to study the effect of dispersal on gut microbial diversity because unlike human and other wild animals studied so far, red squirrels generally do not move around within the

population. In addition, mother raises young in the absence of father, making it possible to dissect host genetics and maternal effects on gut microbiota. We hypothesize that red squirrel gut microbiota are strongly influenced by seasonal diet variation. In addition, we hypothesize that both host genetics and limited dispersal affect gut microbial diversity therefore we expect individuals that are genetically related or live in close proximity will have similar gut microbial profiles.

Results

Red squirrel gut microbiota profile

We analyzed the gut microbial communities of North American red squirrels using 905 fecal samples collected from 363 red squirrels Samples varied in both time (multiple years and seasons) and space (from 6 study grids that were geographically separated from each other), and were from hosts of both sexes, different ages and relatedness. The metadata associated with individuals in the main study grid KL (n=549 samples) were listed in Table 1. After rarefaction to 4000 reads/sample, we sorted high-quality reads into 12,833 operational taxonomic units (OTUs) using a sequence identity cutoff of 97%. On average we detected 575 ± 139 OTUs per sample.

Taxonomic assignment revealed a fairly typical rodent profile (Figure 1A): the dominant phyla were Firmicutes (88.6% of total reads), Bacteroidetes (9.0%), and Proteobacteria (1.7%), with a tail of 10 rare phyla that together accounted for the remaining 0.7% of the reads. The prevalence-abundance distribution of genera showed a “L” shape with a heavy long tail toward the left (Figure 2), indicating that the most abundant taxa were present in almost all samples, while rare taxa accounted for most of the membership difference in gut microbiota of red squirrel population. Specifically, the 10 most abundant genera (5% of the total detected genera) were

each present in more than 97% of the samples and together made up 41.5% of the total reads: *Coprococcus* (abundance: 12.3%, prevalence: 100%), *Blautia* (7.3%, 100%), *Oscillospira* (6.2%, 99.6%), *Clostridium* (3.2%, 98.8%), *Ruminococcus* (2.7%, 99.9%), *Prevotella* (2.6%, 99.9%), *Dorea* (2.0%, 100%), *Anaerostipes* (1.92%, 97.4%), *Bacteroides* (1.87%, 99.4%) and *Faecalibacterium* (1.4%, 99.7%) (Figure 1B). As such, they constituted the "core microbiota" of red squirrel gut microbial community. On the other hand, rare taxa appear to be more sample specific. Among the 189 genera detected, 167 (88.4%) were present in less than 50% of samples. Similarly, 11,618 OTUs (90.5% of total OTUs) appeared in less than 10% of samples. On average, only 56% of genera and 20% of OTUs were shared among samples (average Jaccard distance = 0.44 at the genus level, 0.80 at the OTU level).

Remarkable seasonal variation in gut microbiota diversity and composition

We found remarkable seasonal variation in the gut microbiota composition at the genus level that clearly delineated samples collected in spring (February through April), early summer (May and June) and late summer (July and August) (Figure 1B). Consistently, principle coordinate analysis (PCoA) revealed a clear seasonal pattern in which samples were largely partitioned by season (Figure 3). Canonical correspondence analysis (CCA) confirmed that season explained significant variation of gut microbial composition ($P < 0.001$, Table 2). The seasonal variation was not simply due to the turn over of the host population because longitudinal data collected from the same individual also displayed a seasonal pattern. (Within season Jaccard distance 0.75 < Between season Jaccard distance 0.79, Wilcoxon rank sum test, P <2E-16, Figure 4).

In order to identify key genera that were strongly associated with season, we performed random forest tests. Using genus composition, random forest models were able to differentiate

seasons with an overall accuracy of 83% (Table 3). We identified 6 abundant and 3 rare genera with strong discriminative power. Interestingly, the top discriminatory genera *Oscillospira*, *Coprococcus* and *Clostridium* all belonged to "core" taxa.

Using JTK_cycle, a nonparametric algorithm for detecting rhythmic elements in circadian clock studies, we confirmed that seasonal fluctuations in the abundance of these key genera were repeatable across years (Figure 5). In total, we found 15 genera showing a strong seasonal periodicity of 11-12 months ($P < 0.001$), including all highly discriminatory genera except unclassified Coriobacteriaceae. Many of the 15 genera were also core genera. Among them, *Coprococcus* and *Oscillospira* exhibited the largest periodic fluctuation in relative abundance (amplitude). As shown in Figure 6, there was a clear tradeoff of "core" genera *Oscillospira* and *Coprococcus*, which peaked in early summer and late summer respectively.

The seasonal changes in gut microbiota occurred in parallel with the shift in red squirrel's dietary composition (Figure 7). In spring, red squirrel mainly live on hoarded cones and mushrooms. In early summer, red squirrels add significant amount of fresh spruce buds and needles to their diet. In late summer, red squirrels gradually switch from hoarded cones in their diet to newly available cones. Mantel test showed that diet and gut microbiota compositions were significantly correlated (Bray-Curtis distance, $r = 0.44$, $P=0.003$). To further explore what specific components of the diet correlated with the changes to gut microbial community structure, we analyzed the association of food items with the seasonal rhythmic genera. The elevated level of Oscillospira correlated with increased buds intake in the early summer ($R^2 = 0.36$, $P = 0.007$). In contrast, the relative abundance of *Coprococcus* was positively associated with new cones and snow, and negatively associated with buds ($R^2 = 0.92$, $P = 4.8E\text{-}6$). Percentage of truffle mushroom best predicted the levels of Clostridium in red squirrel gut microbiota ($R^2 =0.25$, $P = 0.03$).

Alpha diversity (measured by Chao1) also displayed a distinct cyclical pattern (JTK_cycle: adjusted P value 4.5E-11, period 12 months, amplitude 101.0). Within each year, species richness reached minima in the early summer and maxima in the late summer (Figure 8). Interestingly, the overall microbial species richness decreased from 2008 to 2010, which coincided with the natural decrease in red squirrel population density over these years (2008: 1.46 squirrels/ha; 2009: 1.15 squirrels/ha; 2010: 0.93 squirrels/ha).

Seasonal OTU co-occurrence network

To investigate how species interactions and the structure of red squirrel gut microbial community changed over time, we reconstructed OTU co-ocurrence network in each season. Analyses of OTU network revealed scale-free network structures in all three seasons (power law, $R^2 > 0.6$). Despite the overall similarity in network structure (Supplementary table 1), the key hub species (species with most connections to other species) varied from season to season, indicating distinct species-species interactions in each season (Figure 9). In spring, a *Coprococcus* species (OTU 21475) was the most dominant hub in the network. In early summer, it faded out of the network and an *Oscillospira* species (OTU 54301) became the most dominant hub. Nevertheless, there was still notable continuity in the network transitions. For example, spring and early summer both had OTU 47644 (Unclassified Ruminococcaceae) as a prominent hub. Early summer and late summer had more hubs in common: OTU 67162 (*Dorea*), OTU 100783 (*Clostridium*) and OTU 32425 (Unclassified Ruminococcaceae). Late summer and spring both had OTU 21475 as their most dominant hub. Interestingly but not surprisingly, most of the hub species belong to the core genera.

Effect of food supplement on red squirrel gut microbiota

In 3 out of the 6 study grids, peanut butter were supplied from October to May in order to

increase the over winter survival rate. The average population density in the grids with food supplement had a two-fold increase compared to the other three grids without supplement (With supplement: 3.13 squirrels/ha; Without: 1.58 squirrels/ha). To assess the effect of food supplement on gut microbiota, we performed PCoA analysis of 225 samples collected in May 2008 from female squirrels in all 6 grids. Samples displayed clear separation by food supplement group (Figure 10A). Grids with food supplement had more *Sutterlla* and *Ruminococcus*, and less *Coprobacillus*, *Clostridium* and *Anaerostipes* (Wilcoxon rank sum test, FDR adjusted $P<0.05$).

Biogeography of red squirrel gut microbiota

To investigate whether red squirrel gut microbiota had any biogeographic structure, we carried out analyses on the same set of samples (n=225) described above. To control for potential confounding effects of food supplement and host relatedness (red squirrels disperse on average 96 m from their mother's territory), we only compare samples within the same food group and from unrelated individuals. Only one sample from each red squirrel was included to avoid artifacts caused by within individual comparison.

In PCoA analysis, samples clustered by grid within each food group (Figure 10B). CCA analyses performed on grids within each food group separately confirmed that grid was a significant predictor of microbiota beta diversity (Table 2). Consistently, between-grid distances were significantly higher than with-grid distances (Jaccard distance: within-grid 0.75, between-grid 0.77, $P < 0.0001$, Wilcoxon rank sum test).

Next we investigated the effect of geographic distance on microbial diversity. Linear regression analysis revealed significant similarity-distance decays for both within and across grids (Figure 11). Every 1km increment in geographic distance resulted in 1.4 % increase in Jaccard distance within grids ($P < 0.001$) and only 0.2% increase between the grids ($P < 0.001$).

Individuality and maternal effect of gut microbiota

Previous studies have shown that family members tend to have more similar gut microbiota than unrelated individuals and increased levels of host relatedness are associated with greater similarity in gut microbial communities.(Lee *et al.*, 2011; Yatsunenko *et al.*, 2012; Tims *et al.*, 2013; Goodrich *et al.*, 2014). We tested the effect of kinship and genetic relatedness on microbial diversity, taking advantage of the comprehensive pedigree information available for the red squirrel population in our study. To eliminate the seasonal and spatial effect, all comparisons were between samples within the same year, season and grid. We analyzed a total of 121 self pairs (relatedness coefficient r = 1, different sampling time), 59 mother–offspring pairs (r = 0.5), 35 father-offspring pairs (r = 0.5), 13 full sibling pairs (r = 0.5), 77 pairs of half-siblings (r = 0.25, maternal: 37, paternal: 40) and 1293 pairs of unrelated individuals (r ~ 0).

We found evidence of individual gut microbiota signatures. For 21 red squirrels, we analyzed multiple samples (range = 9 to 26 samples per individual; time span ranged from 83 days to 828 days). Samples collected from the same individual at different time points were more similar to each other than to other individuals (Figure 12, Wilcoxon rank sum test, FDR adjusted P = 2.5E-18). To further study how each individual's microbiota changed over time, we carried out a time-decay analysis of microbial similarity. Samples from the same individual showed a rapid decrease in similarity within two weeks, and reached plateau after a few months (Figure 4), following a power law function (Jaccard distance = 0.69*Day^(0.023), R^2 = 0.3, P < 2E-16) similar to what has been observed in a previous study in human (Faith *et al.*, 2013).

Interestingly, except for the mother-offspring pair, microbial similarities of all other related pairs (i.e., father-offspring, half siblings and full siblings) were not significantly different from unrelated pairs (mother-offspring, FDR adjusted P =0.01; other related pairs, FDR adjusted P > 0.1, Figure 12), indicating that genetic relatedness of the host did not affect the gut microbial

composition. Consistently, Mantel test did not find any significant correlation between host relatedness and microbial similarity. On the other hand, mother-offspring pairs had significantly higher microbial similarity compared to all other pairs including the father-offspring pairs (Wilcoxon rank sum test, FDR adjusted P = 0.03) (Figure 12).

Relative contribution of environmental and host factors

We next performed PERMANOVA ("adonis") on beta diversity matrices to assess the relative contribution of environmental factors (e.g. year and season) and host factors (e.g sex and age) simultaneously. Overall the results revealed that in wild red squirrel population, environmental factors explained 5~20 times more variations in gut microbial community structure than host factors (Table 4). When bacterial relative abundance was taken into consideration, season and year had more explanatory power (Jaccard: 4.1% by season and 2.6% by year; Bray-Curtis: 10% by season and 5% by year). This is consistent with our observation that the relative abundance of "core microbiota" fluctuated with season and year, but the membership was more stable. While explained additional variations in gut microbiota diversity, the contributions of host factors (sex and age) were relatively small. There remained a high percentage of variations that cannot be captured by our measurements (Jaccard: 92.4%; Bray-Curtis: 83.8%).

Discussion

In this study, we focused on a well-characterized red squirrel population to assess the contribution of environmental and host factors in shaping gut microbiota structure. At the phylum level, red squirrel gut microbiota composition is broadly similar to those of other mammalian gut microbiota (Ley *et al.*, 2008; Muegge *et al.*, 2011; Delsuc *et al.*, 2014), with

Firmicutes (88.6 %) and Bacteroidetes (9.0 %) being the two major phyla. This result is consistent with our present understanding that mammalian gut harbor a highly constrained set of bacterial phyla adapted to the gastrointestinal tract condition (Ley *et al.*, 2008). Unlike wild mice dominated by genus *Lactobacillus* (Carey *et al.*, 2013; Maurice *et al.*, 2015), red squirrels have high level of genera *Coprococcus* (12.3%) and *Oscillospira* (6.2%), but very low level of *Lactobacillus* (0.88%). This might be due to their dietary preference, since red squirrels are herbivorous (more specifically granivorous) and primarily feed on seeds, young leaves and mushrooms, whereas wild mice are omnivorous.

Notably, red squirrels have remarkably low variation in gut bacterial phyla and shared a core set of genera across time (year, season, host ages), space (grid) and family (Figure 2). Such a core set is rarely found in wild animal populations that usually experience large variation in their natural environmental conditions. A likely explanation for this observation is that red squirrels live in the boreal forest region with relatively low biological diversity (Pastor *et al.*, 1996), thus low diversity in their diet.

We found a remarkable seasonal rhythm in red squirrel gut microbial composition, manifested by a tradeoff of the relative abundance of two core genera, *Oscillospira* and *Coprococcus* in early and late summers (Figure 6). Consistently, the network analyses revealed shifts in key hubs in early summer from *Coprococcus* to *Oscillospira*, and a swap in late summer (Figure 9). The seasonal rhythm in gut microbial structure is clearly associated with seasonal dietary changes. The shift in microbiota composition coincides with emergence of fresh spruce buds in early summer and fresh spruce cones in late summer (Figure 7). Accordingly, we found that the level of *Oscillospira* was positively correlated with buds, whereas the level of *Coprococcus* was positively correlated with new cones and snow, and negatively correlated with buds. *Oscillospira* are frequently found in cattle and sheep rumen and increase significantly in

relative abundance when hosts are feeding on fresh forage diet (Mackie *et al.*, 2003), which is consistent with our findings. Species of *Coprococcus* genera are carbohydrates fermenters and producers of butyric and acetic acids (Holdeman and Moore, 1974).

That fact that *Oscillospira* and *Coprococcus* are present in virtually all the samples we surveyed suggests that they were long-term gut residents and not foodborne. Supporting this, the same study found no evidence of *Oscillospira* on fresh forage (pasture grass) fed to cattle or in soil (Mackie *et al.*, 2003). Our result suggests that red squirrel gut microbiota switch between alternative stable states in response to recurring seasonal dietary changes. This may resulted from continuous selective pressure on gut microbial community during host-microbiota coevolution. Gut microbiota adapted to seasonal dietary shift can rapidly shift their metabolic activity, provide the host dietary flexibility, maximize energy extraction and likely increase the fitness of the host-microbe ecosystem.

Biogeographic patterns have been observed in human and house mice populations (Yatsunenko *et al.*, 2012; Linnenbrink *et al.*, 2013; Maurice *et al.*, 2015; Zhang *et al.*, 2015). All patterns were detected between populations living in different countries or continents. We found evidences for a weak but significant spatial structure at a much smaller local scale. Microbial composition varies across six study grids within a few kilometers. Moreover, a similarity-distance decay relationship was found within a population. Distance-decay patterns in microbial communities can be driven by environmental factors that vary across space, as recapitulated by the hypothesis that "everything is everywhere, but the environment selects" (Becking, 1934). Alternatively, the spatial patterns can be due to dispersal limitation, as it allows historical effect to influence contemporary community structure. We have controlled for potential confounding environmental factors in our analysis (we only included samples from a single year, season and sex). Thus we think the most likely explanation for this biogeographic pattern is dispersal

limitation of gut microbes, although we cannot rule out unmeasured spatially structured environmental factors. Red squirrels defend territories vigorously with territorial calls, and have relatively limited direct interactions (Smith 1986) compared to animals in the other studies described above. Thus, it is not surprising that red squirrel gut microbiota might be constrained by stronger dispersal limitation, which could result in spatial structures at a small local scale.

Island biogeography theory (MacArthur and Wilson, 1967) can be useful for understanding the microbial diversity if we view each individual gut as an island. Island theory posits that early colonizers could strongly influence the future community composition. Mother can make a large contribution to the species pool that first colonizes offspring. It has been recently proposed that maternal transmission of gut microbiota is universal in animals (Funkhouser and Bordenstein, 2013) and the effects of maternal transmission can be manifested over several generations (Ley *et al.*, 2005). In our study, we found that gut microbiota of red squirrels were significantly more similar to those of their mother than to those of their father and unrelated individuals. This finding indicates not only that gut microbiota in red squirrel can be maternally transmitted, but also that the maternal effect persists to adulthood. This observation is consistent with the fact that female red squirrels raise offspring without any help from males.

In contrast to findings in human populations (Goodrich *et al.*, 2014), we found no evidence indicating host genetics influence gut microbiota diversity in red squirrels. Since mother and father were equally related to offspring, the genetic relatedness cannot explain higher microbial similarity observed in mother-offspring pairs. In addition, gut microbiota were not significantly different between father-offspring, full sibling, half sibling, and unrelated individual pairs.

Conclusion

In summary, we performed a comprehensive survey of gut microbiota of a well-studied wild red squirrels population. Red squirrels harbor a typical rodent gut microbiota with a stable set of core genera. We discovered a remarkably strong seasonal rhythm in the gut microbial structure mainly associated with seasonal dietary changes, and a subtle but significant biogeographic pattern at a fine local scale indicative of limited gut microbial dispersal. Despite the dominant effect of environmental factors, we found clear signatures of individuality and maternal effect in red squirrel gut microbial communities. However, host genetics does not seem to be a significant driver. Taken together, this study contributes towards a better understanding of the various ecological forces underlying the temporal and spatial patterns of gut microbiota in natural environment.

Findings from free-ranging wild animals could guide our future research focus. Follow up studies using red squirrels in captivity with well-controlled dietary intake, calorie content, and closlely monitored host physiology are necessary to understand the following questions. 1) Is diet driving the observed seasonal cycles in red squirrel gut microbiota? 2) Are compositional changes in gut microbiota causing differences in energy harvesting efficiency and host fitness? Answering these questions could provide insights towards a better understanding of how gut microbiota co-evolve with red squirrels in natural environment.

Material and Methods

Sample collection

Subject description. Study subject were natural population of North American red squirrels in the

southwest Yukon (61° N 138° W) near Kluane National Park. Red squirrels in this area have been continuously monitored by the Kluane Red Squirrel Project since 1987 using a combination of live-trapping and behavioral observations. All squirrels were permanently marked with small metal ear tags and regularly monitored from March to September of each year. Several types of data including identity, sex, body mass, reproductive status, territory ownership and dietary information were collected. In this study, we collected 1,000 fecal samples from 363 individuals that span 3 years and 240 hectares. Samples used in our study were described below. A detailed description of the population can be found in (McAdam *et al.*, 2007).

Study grids. The study area consists of six 40 hectare grids (AG, KL, SU, CH, JO, LL) that are 0.2~7.3 km apart from each other (Figure 13). Samples were collected mainly from two grids (KL: n=618; AG: n=232). Samples collected from the other four grids were used to study the biogeographic structure of microbiota between grids (JO: n=25, SU=25, CH=50, LL=50). In grids AG, LL and JO, peanut butter was provided as food supplement from October to May to experimentally increase the population density (Dantzer *et al.*, 2013). The main study grid (KL) was not manipulated and therefore represents the natural environment for red squirrels.

Sampling years and seasons. The seed of White spruce cones (*Picea glauca*) is the major food resource for red squirrels. White spruce is a masting tree species that produces a super-abundance of cones in some years (mast years) and fewer to no cones in other years. Previous studies have shown that yearly variation in the spruce cone production has large ecological and evolutionary impacts on red squirrels (Boutin *et al.*, 2006; Fletcher *et al.*, 2013; Dantzer *et al.*, 2013). Our samples span from year 2008 to 2010, with 2010 being a mast year. Within each year, samples were collected in three seasons: spring (February through April), early summer (May and June) and late summer (July and August).

Diet information. The study area is located in a boreal forest dominated by white spruce (*Picea glauca*) and willows (*Salix spp*.). Red squirrels feed on the seeds in fresh white spruce cones in the fall and hoard them for consumption over winter (red squirrels do not hibernate) and in the next spring. Squirrels also feed on mushrooms, spruce buds, truffles, berries and a variety of items depending on their seasonal availability (Fletcher *et al.*, 2013). Food items in feeding events were visually identified and recorded each year. Since individuals consume similar foods in roughly similar proportions, all feeding events from 2008 to 2010 were aggregated by months in our study to provide suitable estimates of the composition of diet. Red squirrels' diet includes 14 food categories: (1) seeds of hoarded white spruce cones, (2) seeds of fresh white spruce cones, (3) hypogeous fungi (truffles), (4) spruce buds, (5) spruce needles, (6) spruce bark, (7) willow leaves, (8) willow buds, (9) Aspen leaves, (10) bearberry flowers, (11) white spruce witches broom rust caused by the fungus *Chrysomyxa arctostaphyli*, (12) animal material, (13) snow, and (14) unidentifiable items (Figure 7).

Territory. Adult red squirrels defend exclusive territories around a central larder hoard ("midden") containing cached white spruce cones for over winter survival. A juvenile failing to acquire a territory before its first winter will not survive. On average only 26% of offspring survive to 1 year of age. The location of each animal's midden was recorded and used to estimate the geographic distance between individuals.

Age. The average wild red squirrel lifespan is 5 years (maximal: 10 years). Juveniles usually leave the natal area 70 days after birth, and the mean dispersal distance is 96 ± 94 m from the natal area (Berteaux and Boutin, 2000). Red squirrels reach sexual and reproductive maturity at 1 year old. In this study, we collected samples from individuals 0-6 years of age. The age estimation was accurate to within days.

Sex and reproductive status. Red squirrels are sexually monomorphic in adult body mass, and there is no sex-bias in natal dispersal. During breeding season (March to May), males invade the territories of females and chase them. Females raise young without any help from males. During each capture event, the reproductive status of adult females was determined using nipple status and recorded as non-breeding, pregnant (fetus palpable in abdomen), lactating or weaning.

Pedigree.

Pedigree in grid KL from 2008 to 2010 includes 124 individuals, with 78 known maternal links and 83 known paternal links. Maternity was determined by behavior observation before the emergence of juvenile squirrels from their natal areas. Paternity was assigned based on 16 microsatellite loci using CERVUS 3.0 with >= 95% confidence (detailed in Lane *et al.*, 2008)

Fecal sample collection, DNA extraction and 16S rRNA sequencing

Fecal samples collected from underneath live traps were placed into 1.5 mL vials individually using forceps. Fecal samples collected in the colder months (January-April) were generally frozen upon collection. In the warmer months (May-September), the vials were kept on ice and then transferred to a -20 C freezer within 5 h of collection.

We extracted DNA from fecal samples in a 96-well format using the ZR-96 Fecal DNA Kits (Zymo Research, Orange, CA) following the manufacturer's protocol. The V1-V3 hypervariable regions of the 16S rRNA gene were amplified using two universal primers 27F (5'-AGRGTTTGATCMTGGCTCAG-3') and 534R (5'-TTACCGCGGCTGCTGGCAC-3'). We added a unique 8bp barcode to each primer to tag the samples and used a 50 uL reaction for each PCR amplification by QIAGEN Taq polymerase (Qiagen Inc, CA). PCR conditions consisted of 94°C for 3 min, followed by 25 cycles of 94°C for 30 s, 57°C for 30 s, and 72°C for 60 s, with a final extension of 5 min at 72°C. 16S rRNA amplicons from different samples were pooled in

equal molar ratios, then gel purified and sequenced on an Illumina MiSeq platform using the 300bp paired-end (PE) protocol. All liquid transfer steps were performed on a Biomek NXp liquid handling station (Beckman-Coulter Inc., Fullerton, USA).

Sequence processing, quality control, and OTU classification

We filtered sequence reads by base quality using TRIMMOMATIC 0.32 with settings of LEADING = 3, SLIDINGWINDOW = 10:20, and MINLEN = 50 (Bolger *et al.*, 2014). Paired-end reads passing the quality filter were merged using FLASH (-r 301 -f 447 -s 45 -x 0.05) (Magoč and Salzberg, 2011). The successfully merged reads were assigned to samples by barcodes and processed using the QIIME pipeline (Caporaso *et al.*, 2010). We identified chimeric sequences using usearch (Edgar *et al.*, 2011) implemented in QIIME with both de novo and reference-based detection algorithms. Only those sequences that were flagged as non-chimeras with both detection methods were retained. We then removed non-16S rRNA sequences using hmmsearch (Eddy, 1998) against a custom-made 16S rRNA gene model. The remaining reads were clustered to operational taxonomic units (OTUs) by UCLUST (Edgar, 2010) using an identity threshold of 97%. The most abundant sequence of each OTU was selected as the representative sequence, which was then classified using the RDP classifier (Wang *et al.*, 2007). OTUs belonging to mitochondrion or chloroplast were removed. To remove sequencing effort heterogeneity, samples were rarefied to 4,000 reads per sample. Of our initial set of 1000 samples, three were excluded as outliers because the average distance of each of these three samples from other samples were more than 1.5 times the interquartile range above the higher 75% percentile. During rarefaction, 92 samples were removed due to insufficient number of reads, leaving a final dataset of 905 samples.

Effect of environmental/host factors on gut microbial diversity

We measured alpha diversity by Chao1 index. The Chao1 index estimates species richness based on the number of observed OTUs in each sample using the formula:

$$S_{chao1} = S_{obs} + \frac{n_1(n_1 - 1)}{2(n_2 + 1)}$$

where S_{chao1} is the estimated richness, S_{obs} is the observed number of species, n_1 is the number of singleton taxa (taxa represented by a single read in that community), and n_2 is the number of doubleton taxa. If a sample contains many singletons, it is likely that more undetected OTUs exist, and the Chao 1 index will estimate greater species richness than it would for a sample without rare OTUs. To compare beta diversity among samples, we first excluded any OTUs with less than 5 sequence reads. We then constructed beta diversity matrices from OTU table using four distance metrics: Jaccard, Bray-Curtis, unweighted Unifrac and weighted Unifrac distance (Lozupone and Knight, 2005).

To test the predictors of gut microbial composition, we first performed exploratory Principle Coordinate Analysis (PCoA), followed by hypothesis testing via Canonical Correspondence Analysis (CCA) (M. W. Palmer, 1993). PCoA was performed on all beta diversity matrices in QIIME. CCA was performed on the relative abundance of bacterial taxa in each sample and host associated metadata using the vegan package in R. We carried out CCA analysis at bacterial phylum, genus, and OTU levels. For each test, the best model was selected using the log likelihood criterion, and the significance of each predictor was assessed by 999 permutation tests.

To assess the relative contribution of environmental and host factors to the variation of microbial community, we performed PERMANOVA on Jaccard and Bray Curtis distance matrices using the "adonis" function of the vegan package implemented in R (Oksanen *et al.*,

2007).The percentage of variation explained by each factor was measured using R^2, and the significance (P value) of each factor was obtained by 999 permutation tests.

Identifying bacterial taxa with seasonal rhythms

We used supervised random forest model implemented in QIIME (supervised_learning.py) to identify signature genera in each of the three seasons: spring, early summer and late summer. Random forest model classified each fecal sample into one of three seasons using models built on the relative abundance of each genus. Model accuracy was calculated using the 10-fold cross validation error estimate, which was an approximation of how frequently a sample was misclassified. The discriminatory power of each genus was assessed by comparing the classification accuracy with and without including the genus in the model. Genera that led to more loss of classification accuracy were considered to be more discriminatory.

To test whether bacterial genera identified above had seasonal rhythms, we used a nonparametric test JTK_CYCLE (Hughes *et al.*, 2010). JTK_CYCLE has been used in detecting rhythmic elements in circadian clock studies (Thaiss *et al.*, 2014). We tested seasonal periodicity using a window of 11-12 months. Benjamini-Hochberg procedure was used to control the false discovery rate.

Correlation between microbial composition and diet

To test possible associations between dietary items and rhythmic genera identified above, we constructed linear models on each genus. We began with the full model including all dietary items as the explanatory variables and genus relative abundance as the response variable. Non-significant predictor variables were excluded stepwise from the saturated model using the 'step' command, and the best model with the lowest AIC score was selected. We checked model

assumptions by examining the distribution of residuals and plotting fitted values against residuals. We also performed Mantel tests to evaluate the correlation between the distance matrices built based on dietary item variation and bacterial beta diversity distances. The significance of Mantel's r was assessed with 999 permutations.

Spatial structure of microbial communities

As red squirrel are territorial and have home ranges smaller than study grids, spatial variation in environment within/across grids could affect gut microbial ecology. To study the biogeographic structure of the gut microbial communities, we examined the correlation between host territory geographic distance and gut microbiota distance matrices with Mantel test, with the significance of Mantel's r assessed with 999 permutations. To control for temporal variation, we restricted our analysis to samples collected within the same year and season. Only one sample from each red squirrel was included to avoid artifacts caused by within individual comparison. The decay of bacterial community similarity with respect to geographic distance was plotted within and between grids. The rate of distance-decay of the bacterial communities was calculated as the slope of the linear regression of the beta diversity similarity over the geographic distance.

Effect of Kinship on gut microbiota

To assess the effects of genetic relatedness, we calculated pairwise relatedness from the extensive pedigree data available for red squirrel in Grid KL using the R package pedantics. We then performed Mantel tests to evaluate the correlation between the relatedness matrix and beta diversity distance matrices. To further assess the effect of kinship, we divided pairs into six groups: mother-offspring (relatedness coefficient $r = 0.5$, $n = 59$), father-offspring ($r = 0.5$, $n = 35$), full siblings ($r = 0.5$, $n = 13$), half siblings ($r = 0.25$, $n = 77$) and unrelated ($r \sim 0$, $n = 1293$).

Jaccard distances of each group were compared with non-parametric Kruskal-Wallis tests with post hoc comparisons, and corrected using Benjamini-Hochberg false discovery rate (FDR). To control for temporal variation, we restricted all the above analyses to samples collected within the same year and season.

OTU co-occurrence network

Microbial network of significant co-occurrence and co-exclusion interactions was built using the CoNet 1.1.0 plugin (Faust *et al.*, 2012) in Cytoscape (Smoot *et al.*, 2011). Networks were built for each season separately, and only abundant OTUs (average relative abundance > 0.1%) were used. The analyses were carried out with the following parameters: 1,000 initial top and bottom edges; five similarity measures (Spearman, Pearson, Mutual information; Kullbackleibler, and Bray Curtis); null distribution generated by 1,000 permutations with renormalization; 1,000 iterations for bootstraps. Networks built with different similarity measures were merged using the Simes method (Sarkar and Chang, 2012) and a Benjamini-Hochberg false discovery rate (FDR) cutoff of 0.05. NetworkAnalyzer was used to analyze the topological parameters of the resulting networks (Assenov *et al.*, 2008).

Author contributions

TR and MW conceived the project and designed the experiment. BD and AGM collected the samples. TR performed DNA extraction, PCR, library preparation and data analyses. TR drafted the manuscript. MW revised the manuscript and supervised the work. All authors read and approved the final manuscript.

References

Amato, K.R., Martinez-Mota, R., Righini, N., Raguet-Schofield, M., Corcione, F.P., Marini, E., et al. (2016) Phylogenetic and ecological factors impact the gut microbiota of two Neotropical primate species. *Oecologia* **180**: 717–733.

Becking, L.B. (1934) Geobiologie of Inleiding Tot de Milieukunde [Geobiology or Introduction to the Science of the Environment].

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., et al. (2010) 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.* **107**: 18933–18938.

Berteaux, D. and Boutin, S. (2000) Breeding Dispersal in Female North American Red Squirrels. *Ecology* **81**: 1311–1326.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Boutin, S., Wauters, L.A., McAdam, A.G., Humphries, M.M., Tosi, G., and Dhondt, A.A. (2006) Anticipatory reproduction and population growth in seed predators. *Science* **314**: 1928–1930.

Caporaso, J.G., Kuczynski, J., Stombaugh, J.I., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* **7**: 335–336.

Carey, H.V., Walters, W.A., and Knight, R. (2013) Seasonal restructuring of the ground squirrel gut microbiota over the annual hibernation cycle. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **304**: R33–42.

Chandler, J.A., Lang, J.M., Bhatnagar, S., Eisen, J.A., and Kopp, A. (2011) Bacterial Communities of Diverse Drosophila Species: Ecological Context of a Host–Microbe Model System. *PLOS Genet* **7**: e1002272.

Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.M., Cusack, S., et al. (2012) Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**: 178–184.

Dantzer, B., Newman, A.E.M., Boonstra, R., Palme, R., Boutin, S., Humphries, M.M., and McAdam, A.G. (2013) Density triggers maternal hormones that increase adaptive offspring growth in a wild mammal. *Science* **340**: 1215–1217.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., et al. (2015) Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**: 559–563.

Delsuc, F., Metcalf, J.L., Wegener Parfrey, L., Song, S.J., Gonzalez, A., and Knight, R. (2014) Convergence of gut microbiomes in myrmecophagous mammals. *Mol. Ecol.* **23**: 1301–1317.

Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics* **14**: 755–763.

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

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., et al. (2013) The long-term stability of the human gut microbiota. *Science* **341**: 1237439.

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

Fletcher, Q.E., Landry-Cuerrier, M., Boutin, S., McAdam, A.G., Speakman, J.R., and Humphries, M.M. (2013) Reproductive timing and reliance on hoarded capital resources by lactating red squirrels. *Oecologia* **173**: 1203–1215.

Frank, D.N., Robertson, C.E., Hamm, C.M., Kpadeh, Z., Zhang, T., Chen, H., et al. (2011) Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflammatory Bowel Diseases* **17**: 179–184.

Funkhouser, L.J. and Bordenstein, S.R. (2013) Mom Knows Best: The Universality of Maternal Microbial Transmission. *PLoS Biol* **11**: e1001631–9.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., et al. (2014) Human Genetics Shape the Gut Microbiome. *Cell* **159**: 789–799.

Holdeman, L.V. and Moore, W. (1974) New genus, *Coprococcus*, twelve new species, and emended descriptions of four previously described species of bacteria from human feces. *International Journal of Systematic Bacteriology* **24**: 260–277.

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

Hughes, M.E., Hogenesch, J.B., and Kornacker, K. (2010) JTK_CYCLE: An Efficient Nonparametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets. *J Biol Rhythms* **25**: 372–380.

Khachatryan, Z.A., Ktsoyan, Z.A., Manukyan, G.P., Kelly, D., Ghazaryan, K.A., and Aminov, R.I. (2008) Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS ONE* **3**: e3064.

Lane, J.E., Boutin, S., Gunn, M.R., Slate, J., and Coltman, D.W. (2008) Female multiple mating and paternity in free-ranging North American red squirrels. *Animal Behaviour* **75**: 1927–1937.

Lee, S., Sung, J., Lee, J., and Ko, G. (2011) Comparison of the gut microbiotas of healthy adult twins living in South Korea and the United States. *Appl. Environ. Microbiol.* **77**: 7433–7437.

Lee, S.M., Donaldson, G.P., Mikulski, Z., Boyajian, S., Ley, K., and Mazmanian, S.K. (2014)

Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501**: 426–429.

Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005) Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 11070–11075.

Ley, R.E., Hamady, M., Lozupone, C.A., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., et al. (2008) Evolution of Mammals and Their Gut Microbes. *Science* **320**: 1647–1651.

Linnenbrink, M., Wang, J., Hardouin, E.A., Künzel, S., Metzler, D., and Baines, J.F. (2013) The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Mol. Ecol.* **22**: 1904–1916.

Lozupone, C.A. and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**: 8228–8235.

Lucas, F.S. and Heeb, P. (2005) Environmental factors shape cloacal bacterial assemblages in great tit Parus major and blue tit P. caeruleus nestlings. *Journal of Avian Biology* **36**: 510–516.

MacArthur, R.H. and Wilson, E.O. (1967) The Theory of Island Biogeography Princeton University Press.

Mackie, R.I., Aminov, R.I., Hu, W., Klieve, A.V., Ouwerkerk, D., Sundset, M.A., and Kamagata, Y. (2003) Ecology of uncultivated Oscillospira species in the rumen of cattle, sheep, and reindeer as assessed by microscopy and molecular approaches. *Appl. Environ. Microbiol.* **69**: 6808–6815.

Magoč, T. and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957–2963.

Markle, J.G.M., Frank, D.N., Mortin-Toth, S., Robertson, C.E., Feazel, L.M., Rolle-Kampczyk, U., et al. (2013) Sex Differences in the Gut Microbiome Drive Hormone-Dependent

Regulation of Autoimmunity. *Science* **339**: 1084–1088.

Martínez, I., Stegen, J.C., Maldonado-Gómez, M.X., Eren, A.M., Siba, P.M., Greenhill, A.R., and Walter, J. (2015) The Gut Microbiota of Rural Papua New Guineans: Composition, Diversity Patterns, and Ecological Processes. *Cell Rep* **11**: 527–538.

Maurice, C.F., Knowles, S.C.L., Ladau, J., Pollard, K.S., Fenton, A., Pedersen, A.B., and Turnbaugh, P.J. (2015) Marked seasonal variation in the wild mouse gut microbiota. *The ISME Journal* **9**: 2423–2434.

McAdam, A.G., Boutin, S., Sykes, A.K., and Humphries, M.M. (2007) Life histories of female red squirrels and their contributions to population growth and lifetime fitness. *Écoscience* **14**: 362–369.

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A., et al. (2012) Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PLoS ONE* **7**: e39191.

Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., et al. (2012) Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biology* **13**: R79.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., Gonzalez, A., Fontana, L., et al. (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**: 970–974.

Oksanen, J., Kindt, R., Legendre, P., OHara, B., Simpson, G.L., Solymos, P., et al. (2007) The vegan Package. *Community ecology package* 1–190.

Osawa, R., Blanshard, W.H., and Ocallaghan, P.G. (1993) Microbiological Studies of the Intestinal Microflora of the Koala, Phascolarctos-Cinereus .2. Pap, a Special Maternal Feces Consumed by Juvenile Koalas. *Australian Journal of Zoology* **41**: 611–620.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007) Development of the Human Infant Intestinal Microbiota. *PLoS Biol* **5**: e177.

Palmer, M.W. (1993) Putting Things in Even Better Order: The Advantages of Canonical Correspondence Analysis. *Ecology* **74**: 2215.

Pastor, J., Mladenoff, D.J., and Haila, Y. (1996) Biodiversity and ecosystem processes in boreal regions John Wiley & Sons Ltd.

Perez, P.F., Doré, J., Leclerc, M., Levenez, F., Benyacoub, J., Serrant, P., et al. (2007) Bacterial Imprinting of the Neonatal Immune System: Lessons From Maternal Cells? *Pediatrics* **119**: e724–e732.

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

Rausch, P., Rehman, A., Künzel, S., Häsler, R., Ott, S.J., Schreiber, S., et al. (2011) Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proc. Natl. Acad. Sci. U.S.A.* **108**: 19030–19035.

Rehman, A., Sina, C., Gavrilova, O., Häsler, R., Ott, S., Baines, J.F., et al. (2011) Nod2 is essential for temporal development of intestinal microbial communities. *Gut* **60**: 1354–1362.

Ren, T., Grieneisen, L.E., Alberts, S.C., Archie, E.A., and Wu, M. (2015) Development, diet and dynamism: longitudinal and cross-sectional predictors of gut microbial communities in wild baboons. *Environ Microbiol* **18**: 1312–1325.

Sarkar, S.K. and Chang, C.-K. (2012) The Simes Method for Multiple Hypothesis Testing with Positively Dependent Test Statistics. *Journal of the American Statistical Association* **92**: 1601–1608.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**: 431–432.

Thaiss, C.A., Zeevi, D., Levy, M., Zilberman-Schapira, G., Suez, J., Tengeler, A.C., et al. (2014) Transkingdom Control of Microbiota Diurnal Oscillations Promotes Metabolic Homeostasis. *Cell* **159**: 514–529.

Tims, S., Derom, C., Jonkers, D.M., Vlietinck, R., Saris, W.H., Kleerebezem, M., et al. (2013) Microbiota conservation and BMI signatures in adult monozygotic twins. *The ISME Journal* **7**: 707–717.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., et al. (2009) A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484.

Wacklin, P., Mäkivuokko, H., Alakulppi, N., Nikkilä, J., Tenkanen, H., Räbinä, J., et al. (2011) Secretor Genotype (FUT2 gene) Is Strongly Associated with the Composition of Bifidobacteria in the Human Intestine. *PLoS ONE* **6**: e20113.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**: 5261–5267.

Xue, Z., Zhang, W., Wang, L., Hou, R., Zhang, M., Fei, L., et al. (2015) The bamboo-eating giant panda harbors a carnivore-like gut microbiota, with excessive seasonal variations. *mBio* **6**: e00022–15.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., et al. (2012) Human gut microbiome viewed across age and geography. *Nature* **486**: 222–227.

Zhang, J., Guo, Z., Xue, Z., Sun, Z., Zhang, M., Wang, L., et al. (2015) A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. *The ISME Journal* 1–12.

Tables and figures

Table 1. Characteristics of red squirrel in main study grid KL (*n*= 549).

Characteristics	Number of samples
Sex	
Male	220
Female	329
Year	
2008	240
2009	120
2010	189
Season	
Spring	233
Early summer	204
Late summer	112
Age	
0	15
1	166
2	110
3	162
4	57
5	37
6	2
Pedigree information	
Both dam and sire known	248
Dam or Sire known	116
Unknown	185

Table 2. CCA analysis of environment and host traits that predicted variation in gut microbial community composition.

Dataset	Number of samples	Factors tested	Significant factors
Grid KL	549	age, season, sex, year	season ($P<0.001$), year ($P<0.001$), sex ($P<0.001$)
Grid KL (female with reproductive status)	284	age, season, year, reproductive status	season ($P<0.001$), year ($P<0.001$)
Grid with food supplement (AG, LL, JO; 2008, female, early summer)	112	age, grid	grid ($P<0.001$)
Grid without food supplement (KL, SU, CH; 2008, female, early summer)	113	age, grid	grid ($P<0.003$)

Table 3. Highly discriminative genera for predicting seasons by random forest test.

Genera	Mean % increase in error on removal (± SD)	Relative abundance in Spring/Early Summer/Late Summer
Oscillospira	21.88 (± 1.22)	2.6%/15.4%/1.4%
Butyricicoccus	11.97 (± 1.33)	0.12%/0.29%/0.03%
Coprococcus	11.92 (± 1.04)	13.0%/6.3%/22.8%
Unclassified Ruminococcaceae	10.40 (± 0.81)	10.3%/9.2%/7.5%
Clostridium	10.26 (± 1.22)	1.9%/5.4%/4.0%
Unclassified Coriobacteriaceae	5.1 (± 0.76)	0.015%/0.05%/0.004%
Faecalibacterium	4.89 (± 0.45)	1.1%/2.0%/1.0%
Ruminococcus	4.44 (± 0.54)	0.7%/1.3%/0.6%
Parabacteroides	4.02 (± 0.49)	0.18%/0.05%/0.14%

Table 4. Percentage variation in beta diversity explained by environmental and host factors for the KL dataset (n = 549 samples). Jaccard distance was calculated based on OTU membership, and Bray-Curtis distance was calculated based on OTU relative abundance.

Beta diversity	Factors	% Variance explained	P value
Jaccard	Year	2.6	0.001
	Season	4.1	0.001
	Age	0.3	0.001
	Sex	0.4	0.001
Bray-Curtis	Year	5	0.001
	Season	10	0.001
	Age	0.3	0.003
	Sex	0.9	0.001

Figure 1. Bacterial composition of 553 red squirrel fecal samples from grid KL that span year 2008 to 2010. Each column represents one sample. Y-axis values represent the relative abundance of each bacterial taxon. Samples are sorted by the sampling time. (A) Phylum level (B) Genus level.

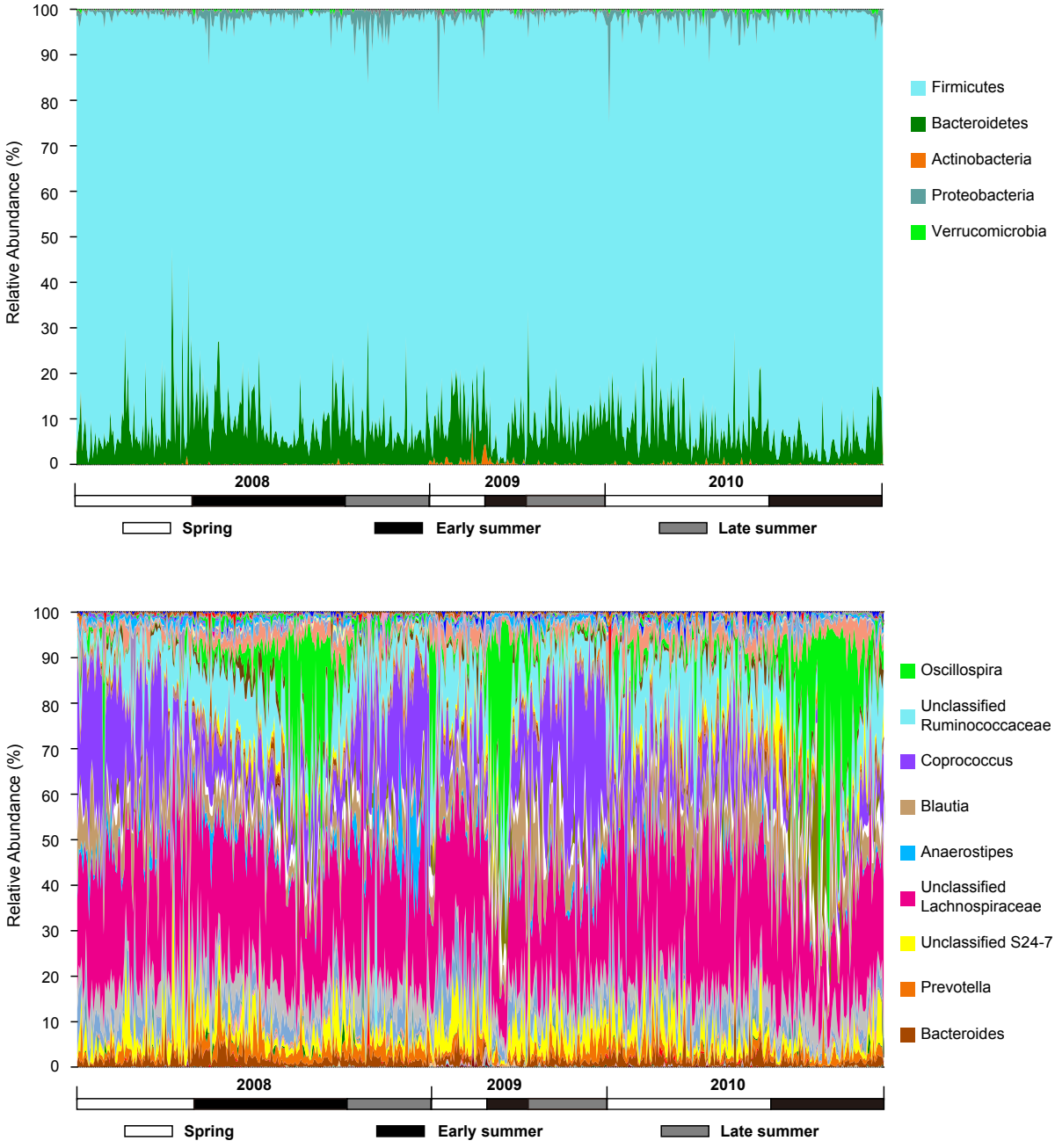


Figure 2. Relative abundance and prevalence of bacterial genera in red squirrel microbiota. The top ten most abundant genera are labeled with their genus names.

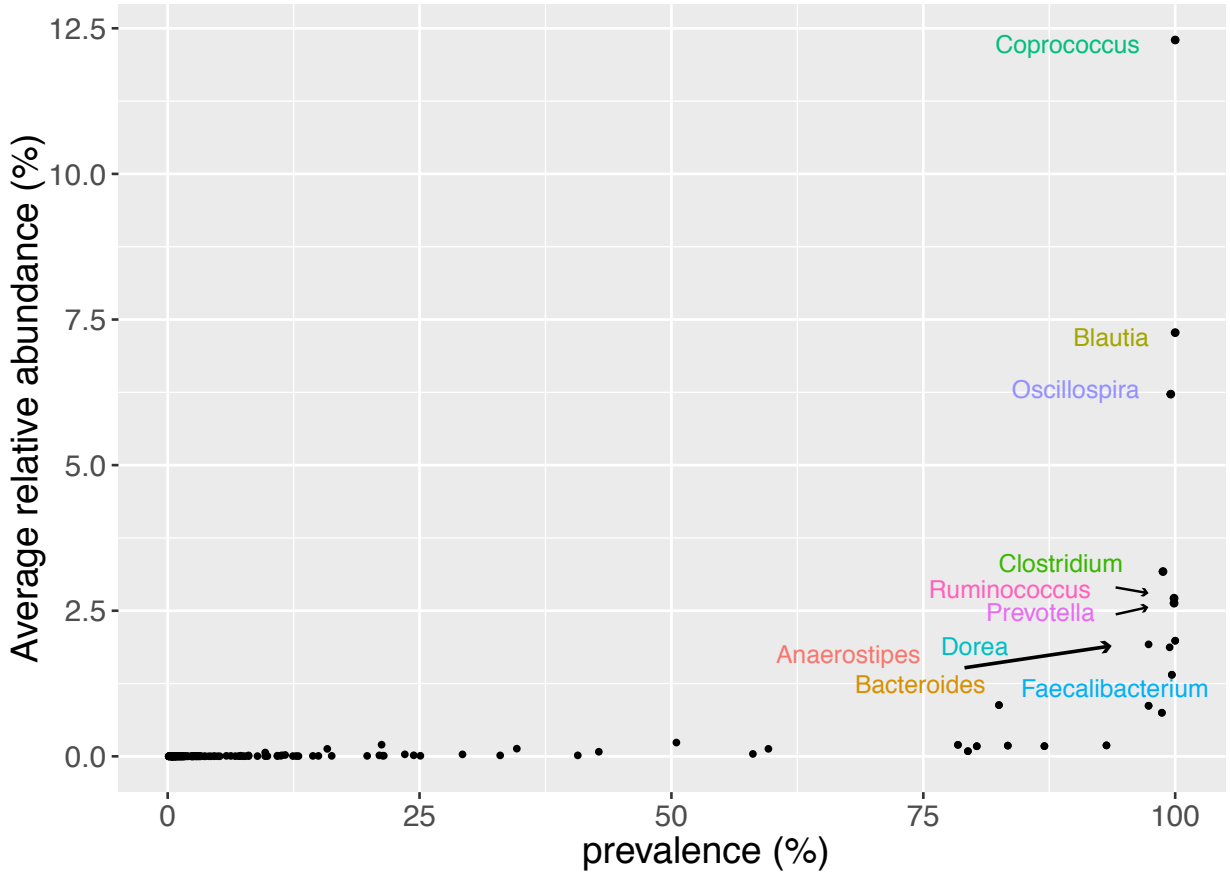


Figure 3. Principal coordinate analyses (PCoA) of red squirrel gut microbial communities in grid KL based on Bray-Curtis distance. Samples are colored by sampling season. The percentage of the variation explained by the first three coordinates are indicated on the axes.

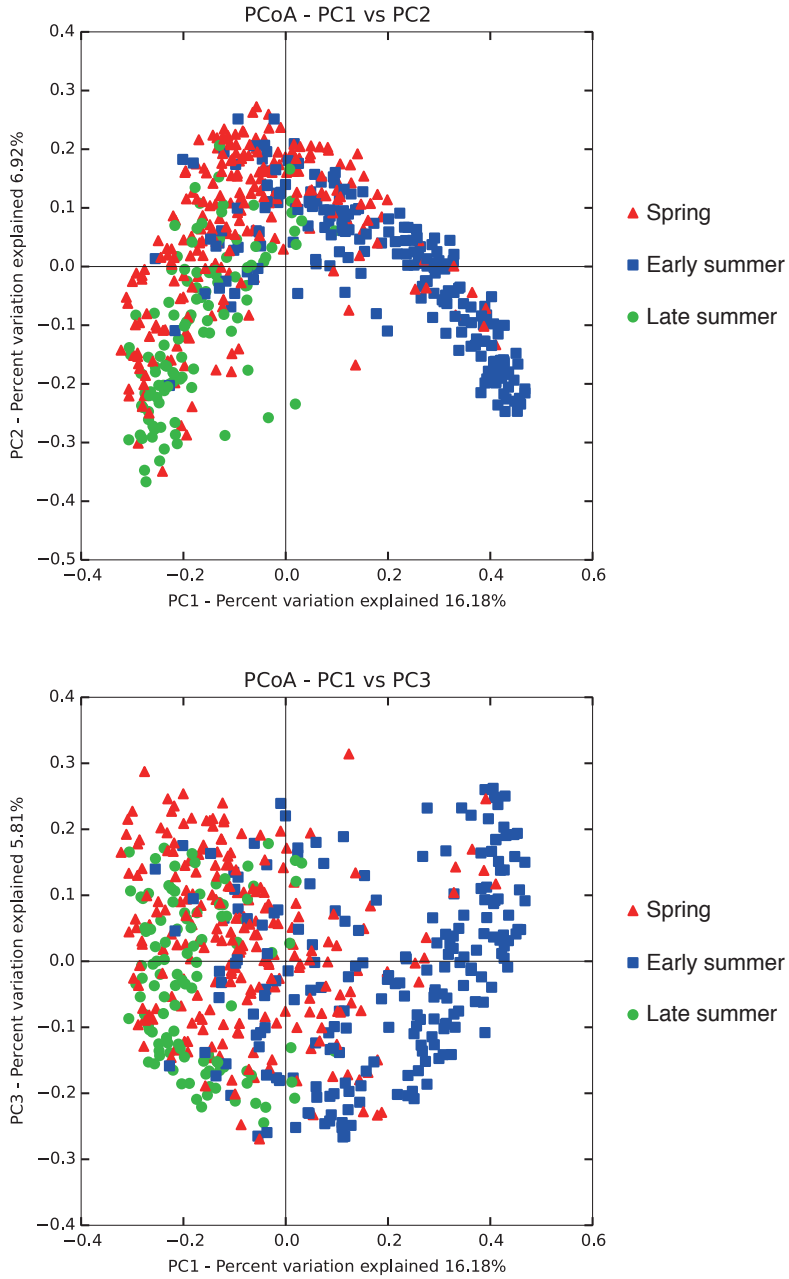


Figure 4. Time-decay of the red squirrel gut microbial communities. Each dot represents a comparison between two samples of the same individual collected at different time points. The colors of dots represent the combination of seasons when the two samples were collected. Y-axis represents the microbiota similarity. The similarity decay as a function of time best fits a power law (blue line). The shade shows the 95% confidence bounds.

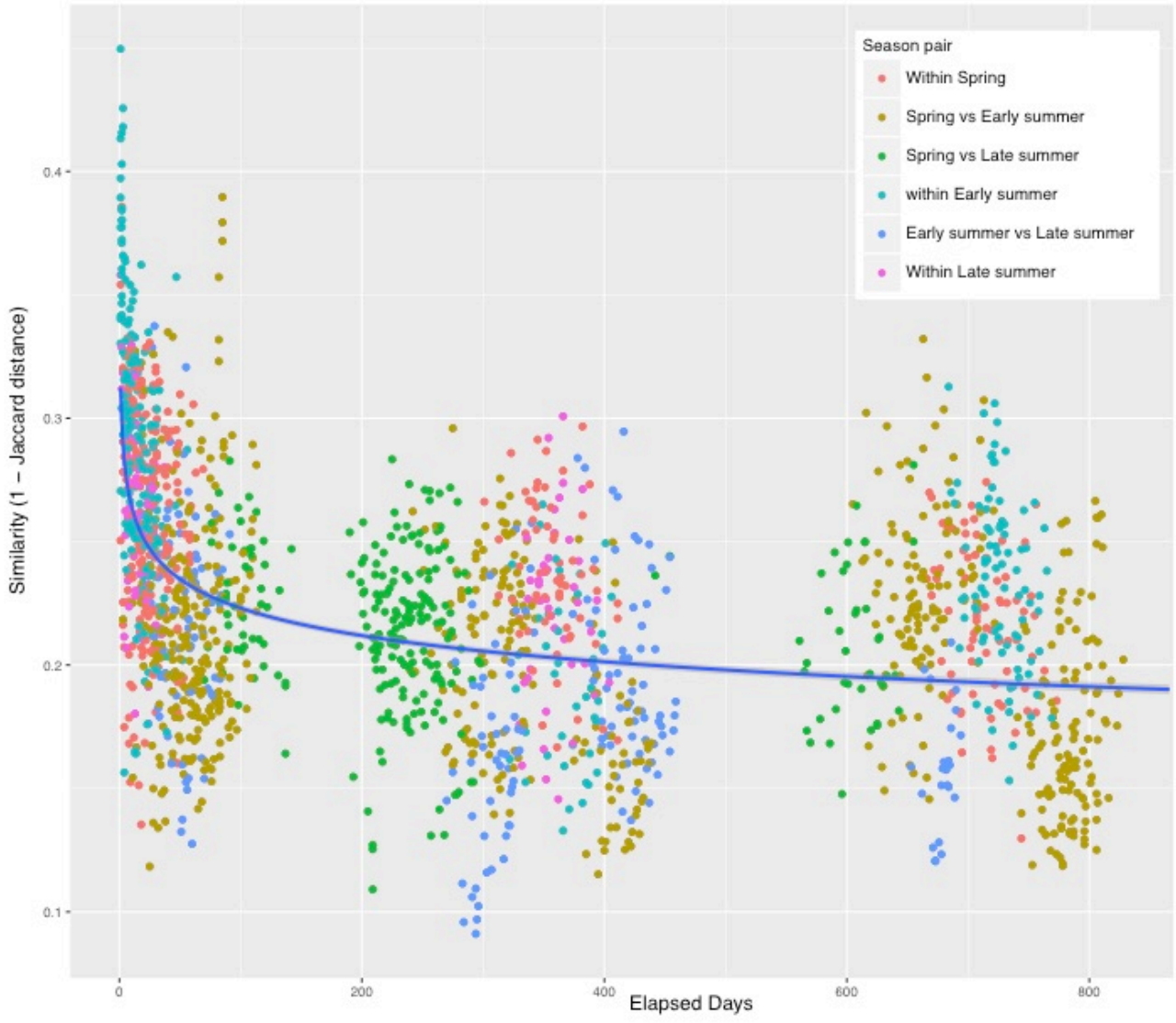


Figure 5. Bacterial genera showing strong seasonal rhythms. X-axis indicates the fluctuation amplitudes and Y-axis indicates the statistical significance of the rhythm. The size of dot represents the average relative abundance of each genus.

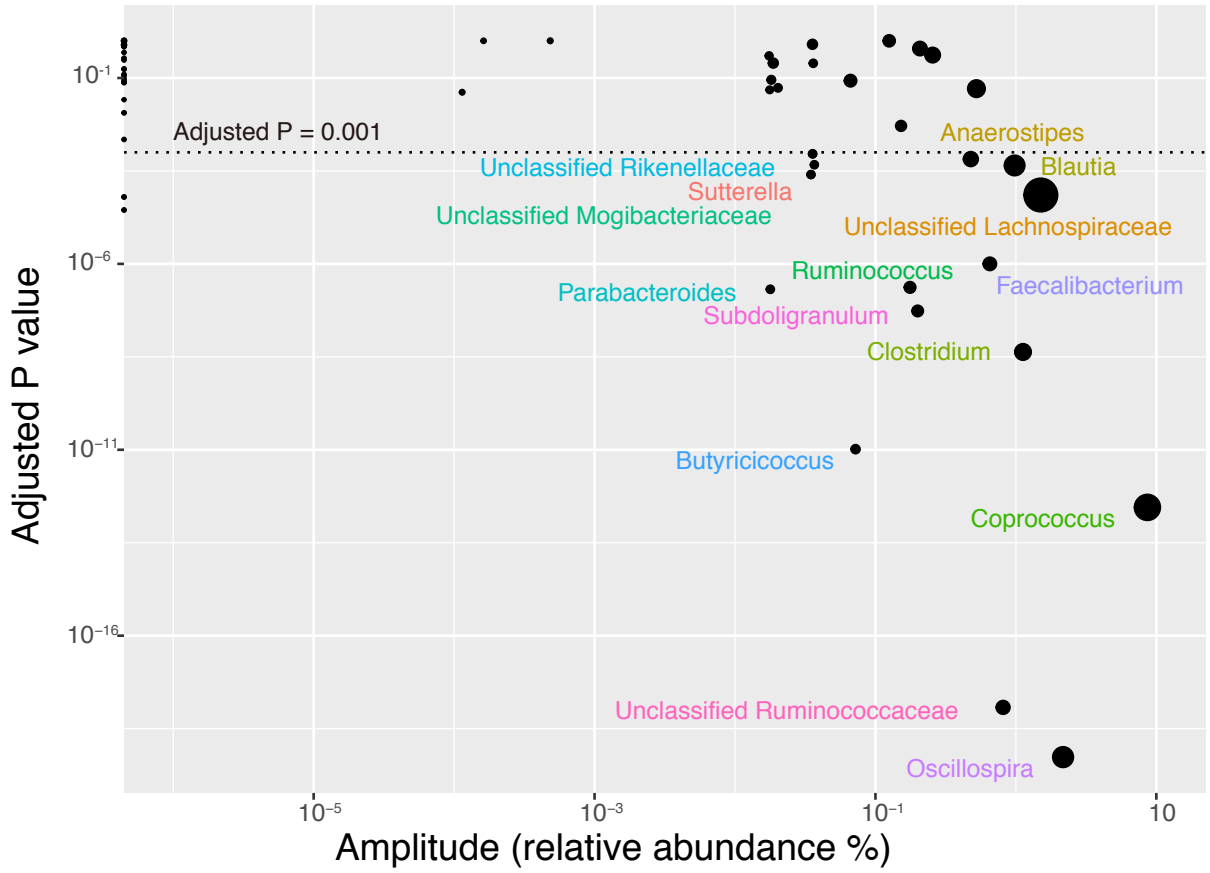


Figure 6. Oscillation of two core genera Coprococcus and Oscillospira over time.

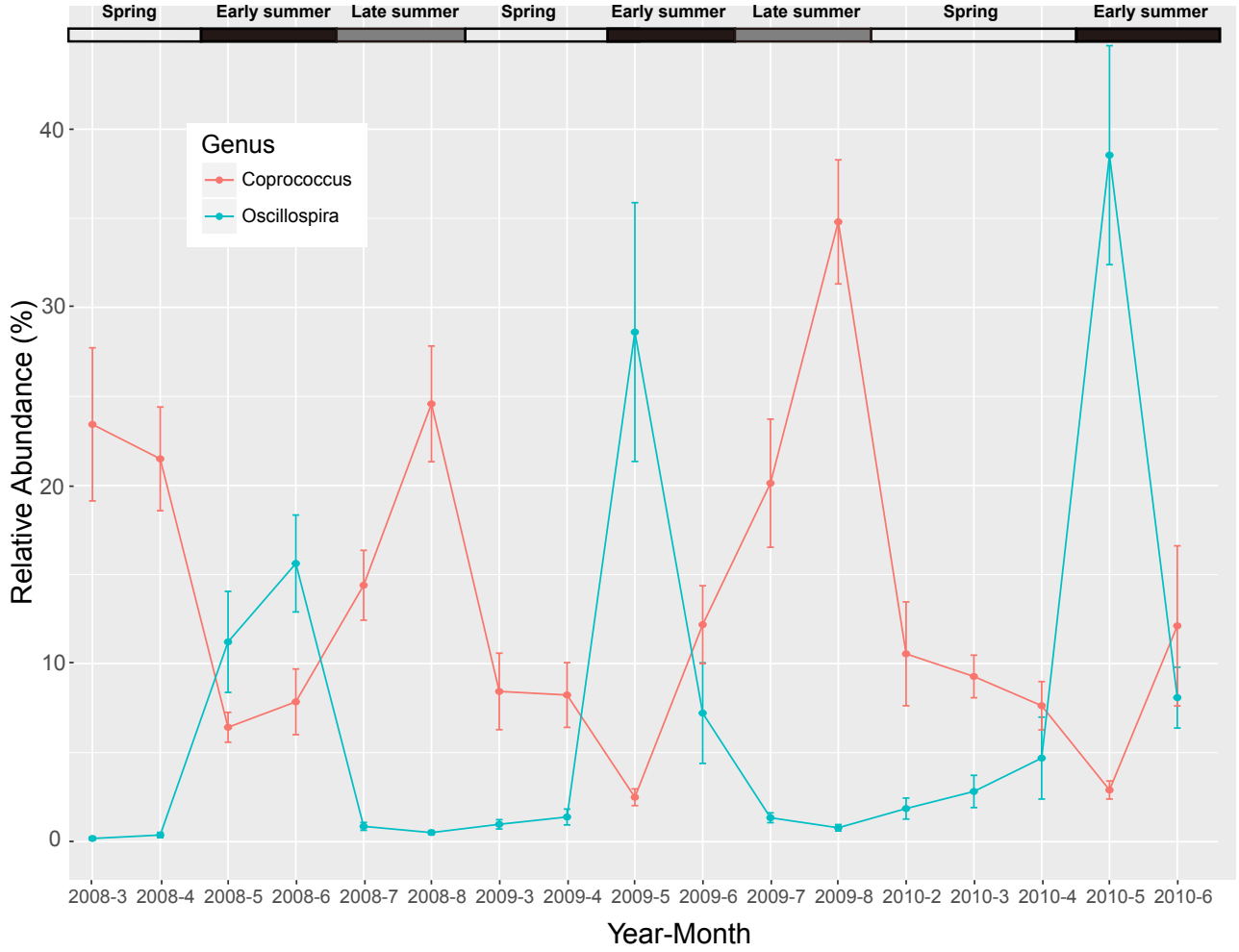


Figure 7. The composition of red squirrel diet across three years aggregated by month. Each color represents a dietary item.

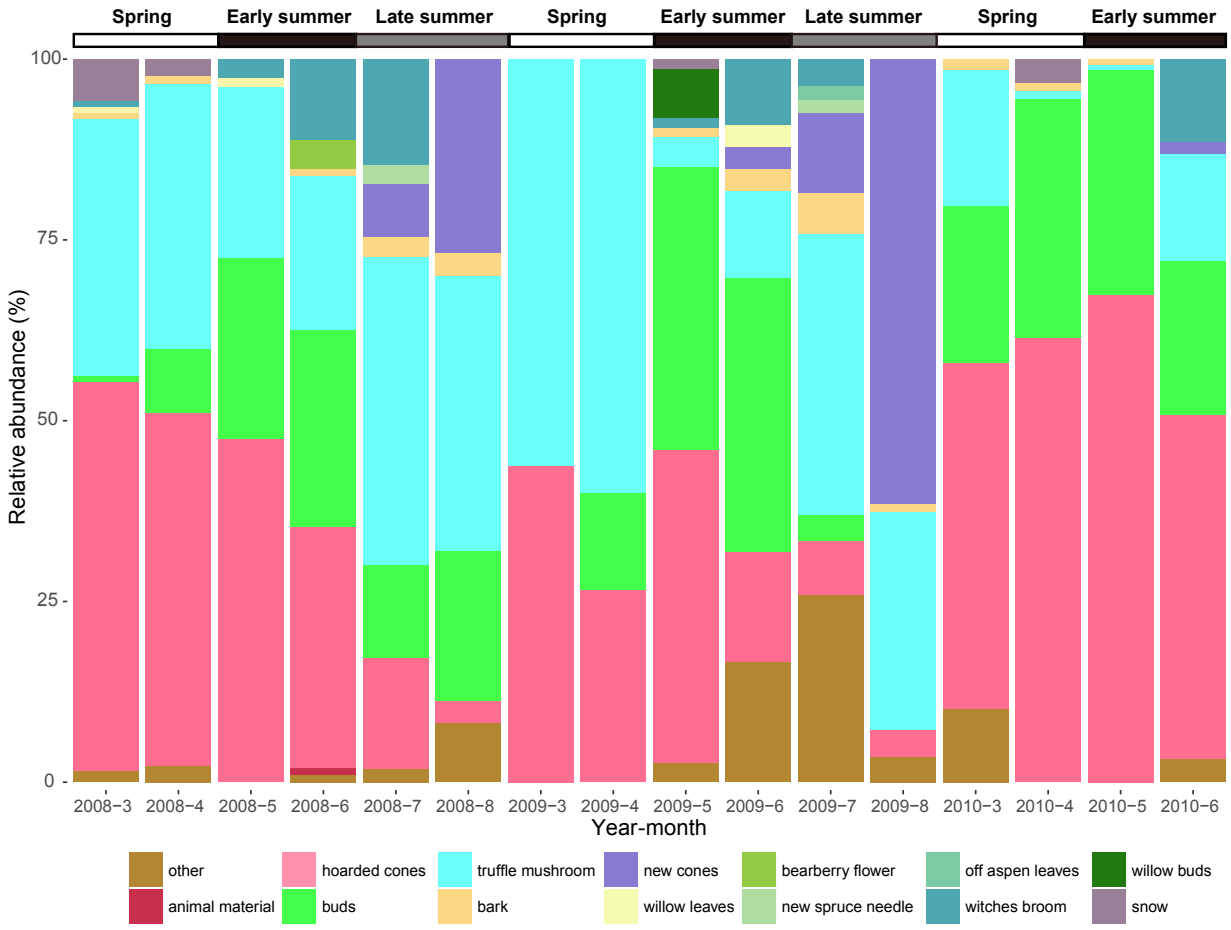


Figure 8. Seasonal rhythm in the alpha diversity of red squirrel gut microbiota. Species richness is estimated by Chao1 index.

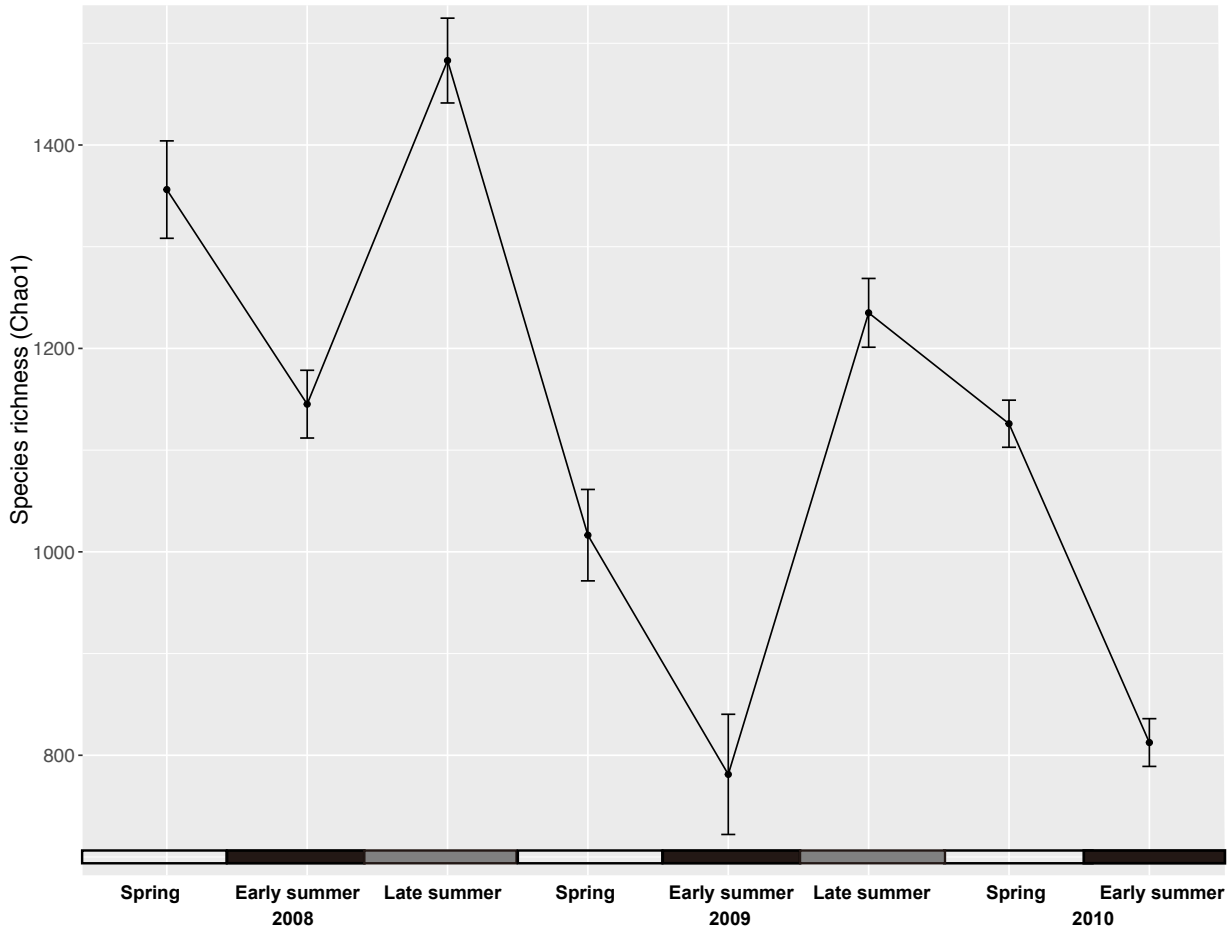


Figure 9. Key hub species in OTU co-occurrence network vary by season. The co-occurrence network is displayed using Cytoscape with the Prefuse Force Directed (edge betweenness) layout. Negative correlations are represented by red edges and positive correlations by green. Each node represents an OTU with >0.1% relative abundance and is colored by bacterial family to which it belongs. Key hub OTUs are labeled with their IDs, genus names and the numbers of positive and negative edges.

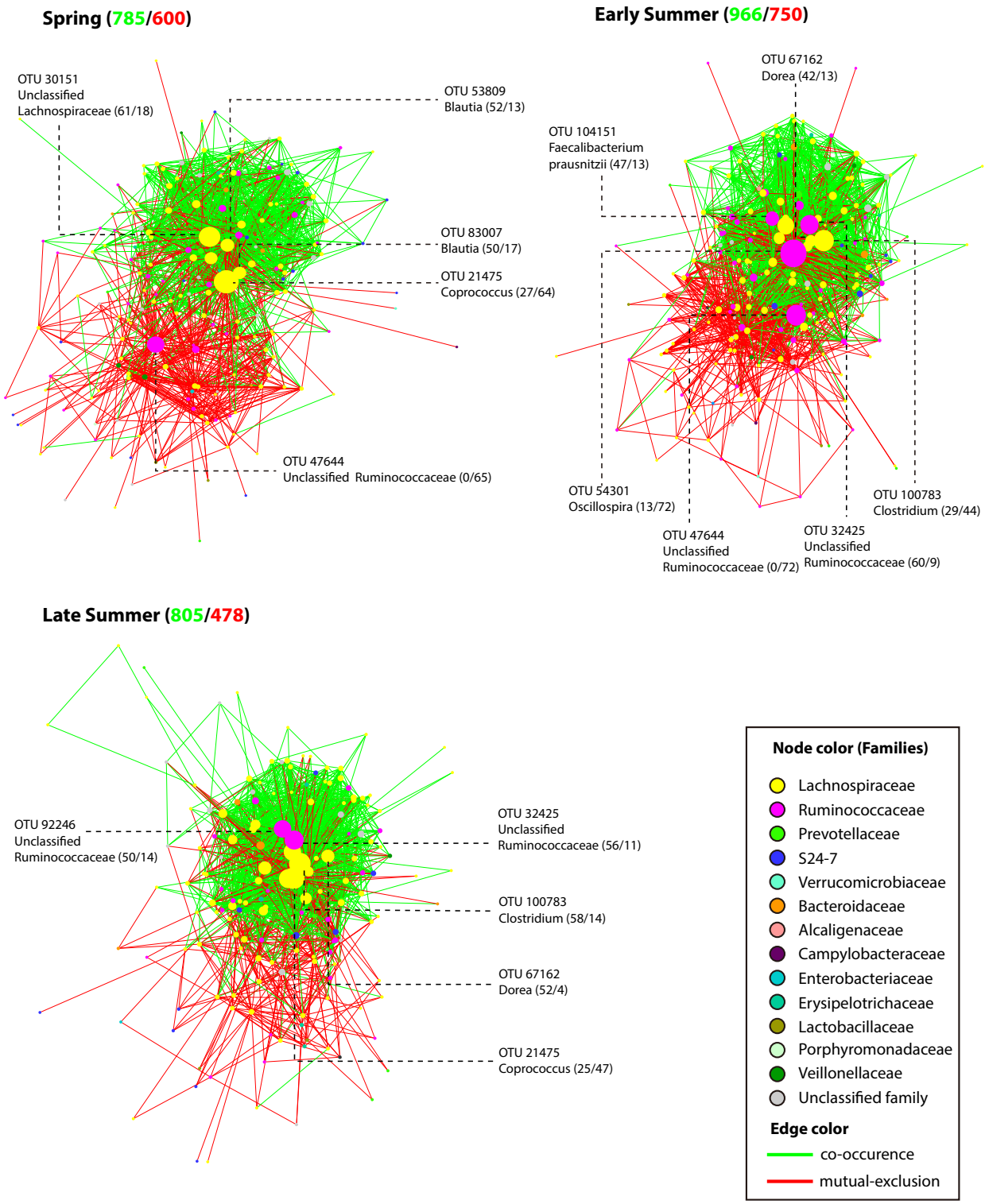

Spring (785/600)
OTU 30151
Unclassified
Lachnospiraceae (61/18)
OTU 53809
Blautia (52/13)
OTU 83007
Blautia (50/17)
OTU 21475
Coprococcus (27/64)
OTU 47644
Unclassified Ruminococcaceae (0/65)
Early Summer (966/750)
OTU 67162
Dorea (42/13)
OTU 104151
Faecalibacterium
prausnitzii (47/13)
OTU 54301
Oscillospira (13/72)
OTU 47644
Unclassified
Ruminococcaceae (0/72)
OTU 32425
Unclassified
Ruminococcaceae (60/9)
OTU 100783
Clostridium (29/44)
Late Summer (805/478)
OTU 92246
Unclassified
Ruminococcaceae (50/14)
OTU 32425
Unclassified
Ruminococcaceae (56/11)
OTU 100783
Clostridium (58/14)
OTU 67162
Dorea (52/4)
OTU 21475
Coprococcus (25/47)
Node color (Families)
Lachnospiraceae
Ruminococcaceae
Prevotellaceae
S24-7
Verrucomicrobiaceae
Bacteroidaceae
Alcaligenaceae
Campylobacteraceae
Enterobacteriaceae
Erysipelotrichaceae
Lactobacillaceae
Porphyromonadaceae
Veillonellaceae
Unclassified family
Edge color
co-occurence
mutual-exclusion

Figure 10. Principal coordinate analyses (PCoA) of red squirrel gut microbial communities across 6 grids based on Jaccard distance. Samples are colored by (A) food supplement status (B) grids. The percentage of the variation explained by the first three coordinates are indicated on the axes.

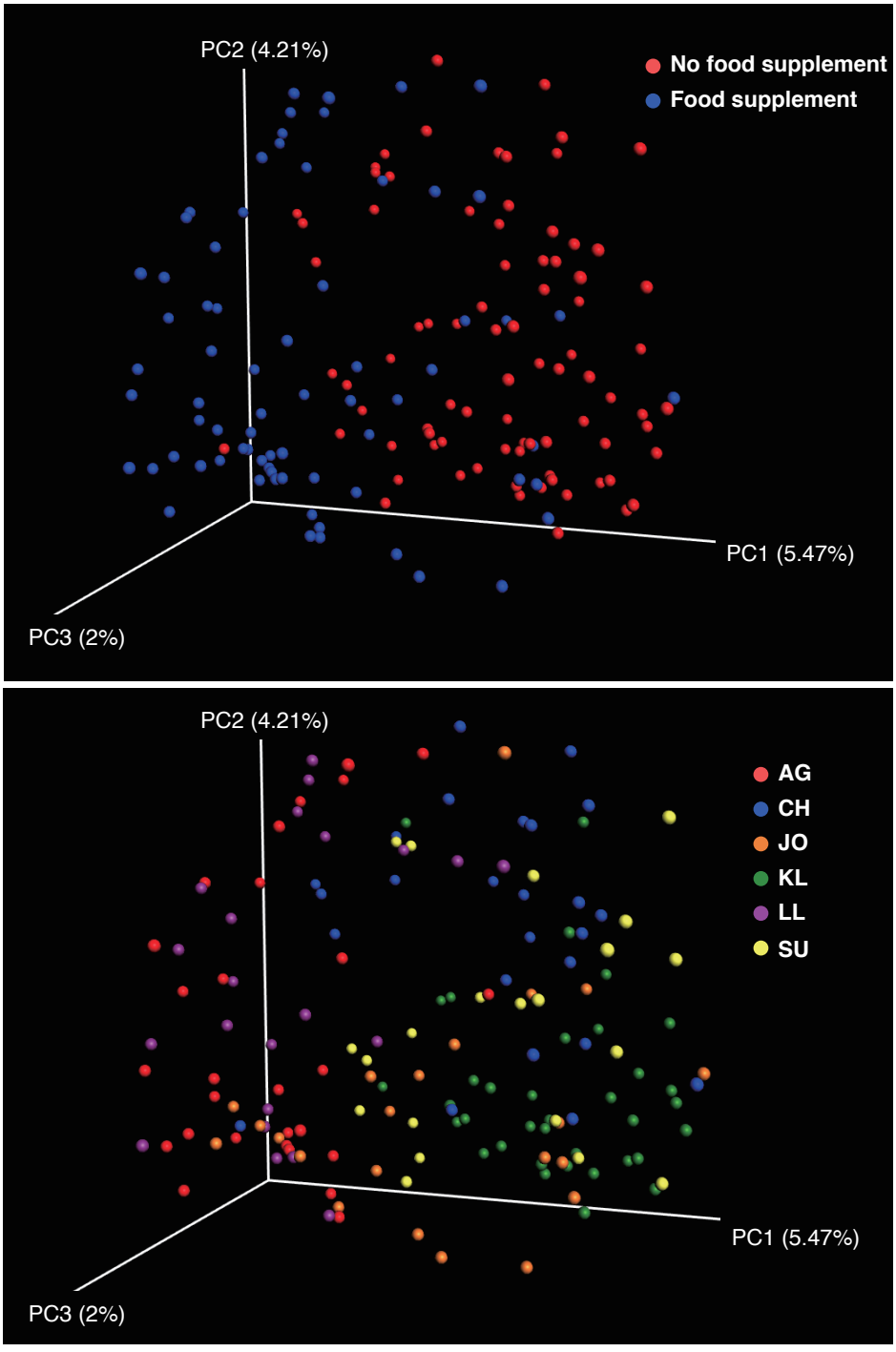


Figure 11. Distance-decay of the red squirrel gut microbial communities within and between grids. Each dot represents a comparison between samples collected at different geographic locations. Y-axis represents the microbiota similarity. The lines denote the linear regressions of microbial similarity over the geographic distance.

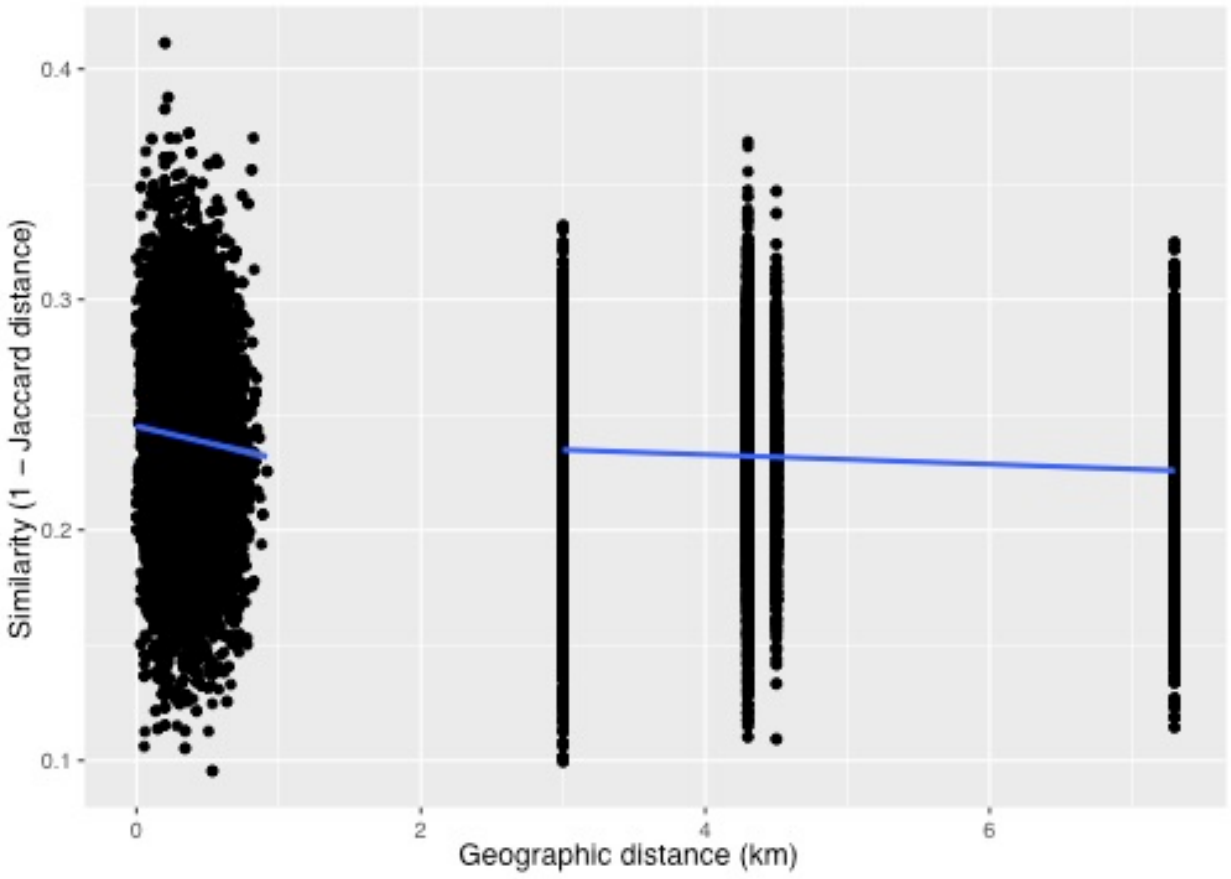


Figure 12. Red squirrel gut microbiota exhibits individuality and maternal effect. Box-and-whisker plots show pairwise Jaccard distances within each relationship groups. Significance values are from non-parametric Kruskal-Wallis tests (FDR adjusted). * P<0.05, ** P<0.01, *** P<0.001.

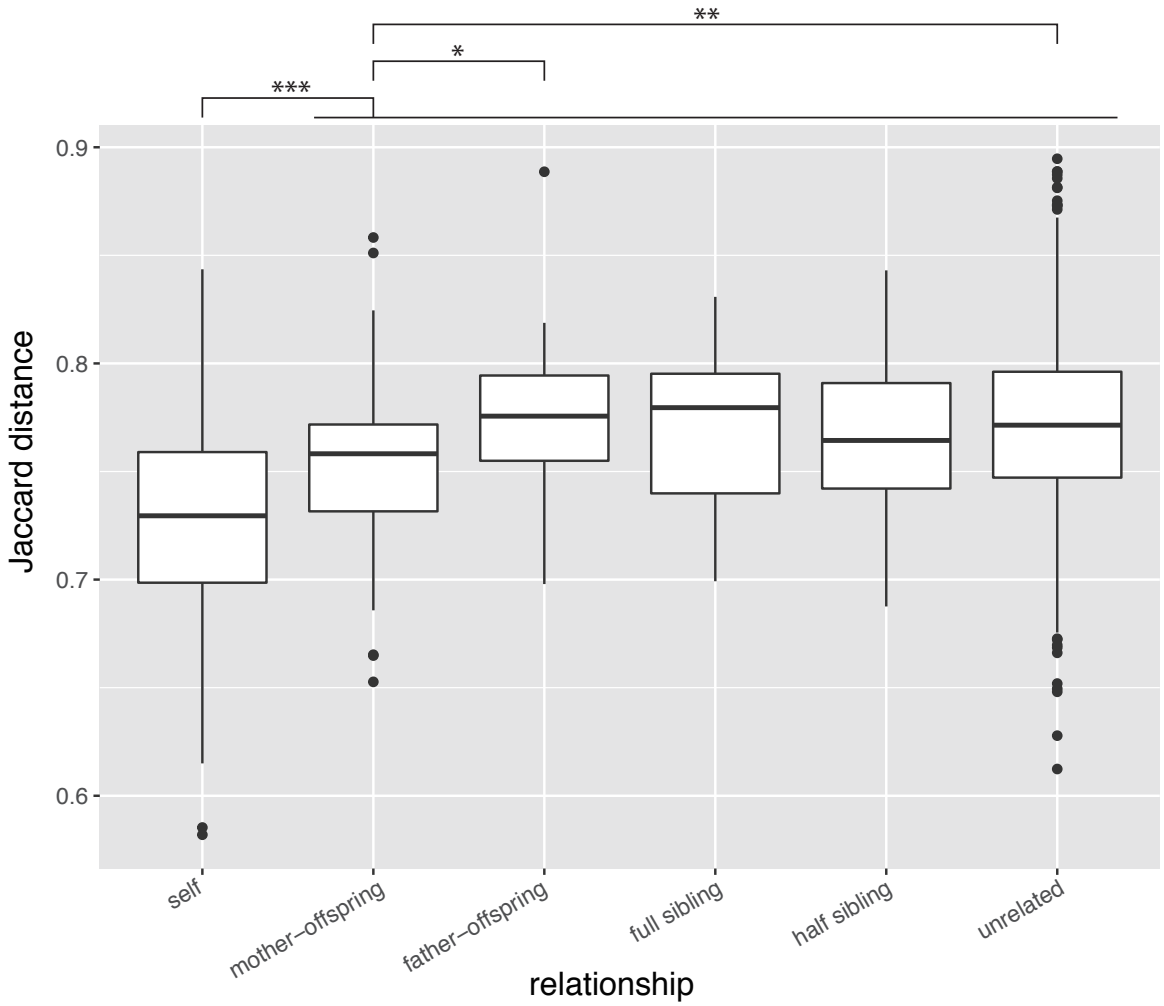
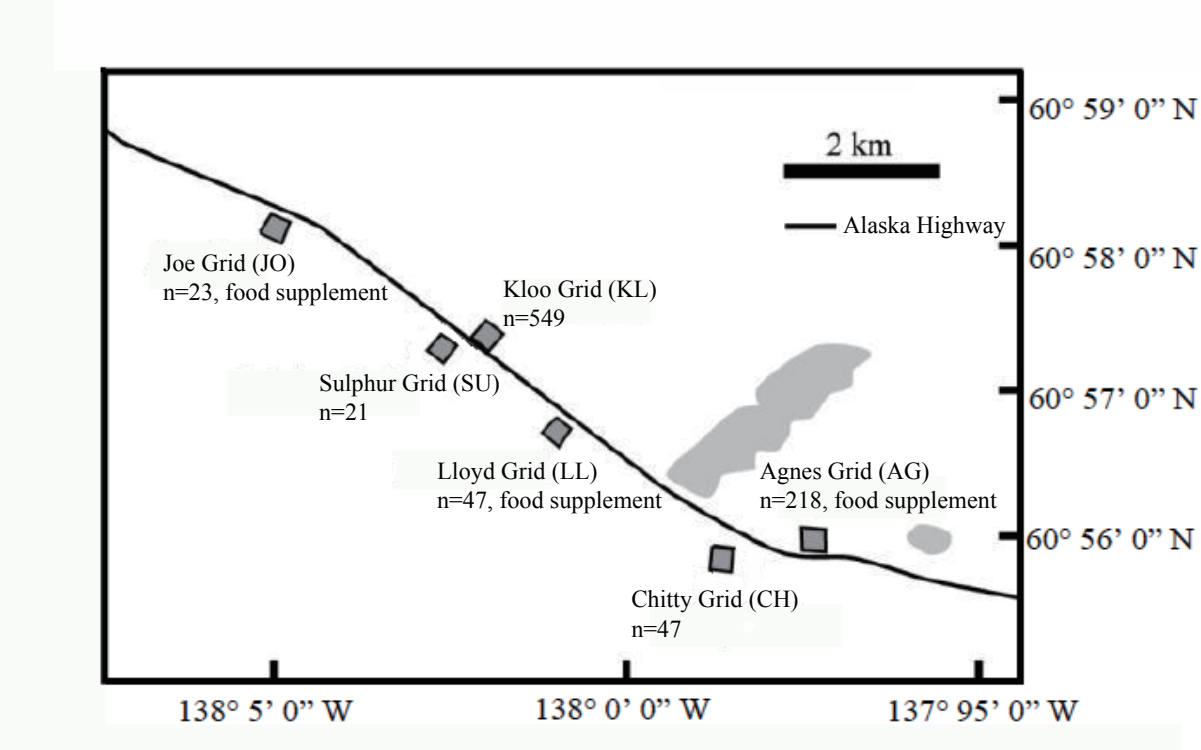


Figure 13. Locations of study grids along the Alaska Highway in Southwest Yukon (61° N 138° W) near Kluane National Park. (Adopted from Villette, 2013) Each grid is labeled with the number of samples collected and the food supplement status.



Supplementary information

Supplementary Table S1. Properties of OTU networks in three seasons.

Network properties	**Spring**	**Early summer**	**Late summer**
Clustering coefficient	0.354	0.355	0.446
Number of nodes	189	188	162
Number of edges (positive/negative)	1385 (785/600)	1716 (966/750)	1283 (805/478)
Network density	0.078	0.098	0.098
Network heterogeneity	0.967	0.889	0.975
Network diameter	5	4	5
Network radius	3	3	3
Network centralization	0.41	0.361	0.353
Characteristic path length	2.316	2.257	2.297
Average number of neighbors	14.656	18.255	15.84
Shortest paths	35532 (100%)	35156 (100%)	26082 (100%)

Chapter 4. Does adaptive radiation of a host lineage promote ecological diversity of its bacterial communities? A test using gut microbiota of Anolis lizards

Formatted as a co-authored manuscript: Ren T, Kahrl AF, Wu M and Cox RM. *Molecular Ecology* (in press)

Abstract

Adaptive radiations provide unique opportunities to test whether and how recent ecological and evolutionary diversification of host species structures the composition of entire bacterial communities. We used 16S rRNA gene sequencing of fecal samples to test for differences in the gut microbiota of six species of Puerto Rican *Anolis* lizards characterized by the evolution of distinct "ecomorphs" related to differences in habitat use. We found substantial variation in the composition of the microbiota within each species and ecomorph (trunk-crown, trunk-ground, grass-bush), but no differences in bacterial alpha diversity among species or ecomorphs. Beta diversity analyses revealed subtle but significant differences in bacterial composition related to host phylogeny and species, but these differences were not consistently associated with *Anolis* ecomorph. Comparison of a trunk-ground species from this clade (*A. cristatellus*) with a distantly related member of the same ecomorph class (*A. sagrei*) where the two species have been introduced and are now sympatric in Florida revealed pronounced differences in the alpha and beta diversity of their microbiota despite their ecological similarity. Comparisons of these populations with allopatric conspecifics also revealed geographic differences in bacterial alpha and beta diversity within each species. Finally, we observed high intra-individual variation over time and strong effects of a simplified laboratory diet on the microbiota of *A. sagrei.* Collectively, our results indicate that bacterial communities are only weakly shaped by the diversification of their lizard hosts due to the strikingly high levels of bacterial diversity and variation observed within *Anolis* species.

Introduction

Gut microbiota are essential to the health and development of their hosts (Turnbaugh *et al.* 2006; Qin *et al.* 2010; Hooper *et al.* 2012), yet we know relatively little about the ecological and evolutionary processes that structure these important bacterial communities. Recent comparative studies show that the similarity of gut microbiota among host species often mirrors the host phylogeny, suggesting that hosts and their microbiota coevolve together (Ley et al. 2008; Degnan et al. 2012; Amato 2013; Sanders et al. 2014). This association between host phylogeny and microbiota could be due to vertical transmission of bacteria, as observed in the case of co-diversification of *Helicobacter pylori* with humans (Falush et al. 2003), as well as phylogenetic conservatism in factors such as diet and habitat, which can strongly influence gut microbial communities (Degnan et al. 2012; Sanders et al. 2014). Adaptive radiations may provide an informative framework in which to simultaneously explore the ecological and evolutionary factors that shape bacterial communities because they are comprised of closely related host species that have diverged to fill different ecological niches.

Lizards in the genus *Anolis* represent a classic example of adaptive radiation, having diversified into nearly 400 morphologically and ecologically diverse species that occur throughout Central and South America and islands of the Caribbean (Losos 2009). On the four major islands of the Greater Antilles (Cuba, Jamaica, Hispaniola, and Puerto Rico), a key feature of this radiation is the convergent evolution of the same 'ecomorphs' (morphologically and behaviorally similar species that share similar ecological niches, but are not necessarily close phyletically; Williams 1972; 1983) across different islands (Losos et al. 1998; Losos 2009). These ecomorphs are classified according to their partitioning of the spatial habitat, and each of the four Greater Antillean islands contains independently derived representatives from most of

the six major ecomorph classes: crown-giant, trunk-crown, trunk, trunk-ground, twig, and grass-bush (Fig. 1A). Although most anoles are insectivorous dietary generalists, ecomorphs may differ in foraging mode (Losos 2009), and partitioning of trophic resources has been observed both within and among sympatric *Anolis* species (Schoener 1967; 1968; Stamps *et al.* 1997). This, along with the fundamental differences in habitat use that characterize ecomorphs, suggests that anoles may provide an intriguing test of the extent to which the evolutionary and ecological diversification of a host lineage structures the biodiversity of entire bacterial communities.

To explore this idea, we sequenced bacterial 16S rRNA genes from anole fecal samples to compare microbiota across six sympatric *Anolis* species on Puerto Rico. These species are members of a single clade that radiated on Puerto Rico and consist of two sister species classified as trunk-crown ecomorphs (*A. evermanni, A. stratulus*), two sister species classified as grass-bush ecomorphs (*A. krugi, A. pulchellus*), and two species classified as trunk-ground ecomorphs (*A. cristatellus*, *A. gundlachi*, Fig. 1B). Hence, we predicted that bacterial communities would be more similar within than among these three ecomorph pairs, a difference that could reflect the phylogenetic affinity and/or ecological similarity of each pair. Next, to test whether phylogenetically distant but ecologically similar species differ in their microbiota when compared in sympatry, we capitalized on the invasion of southern Florida by two trunk-ground ecomorphs from different sources in the Greater Antilles: *A. sagrei*, which has been repeatedly introduced to southern Florida from Cuba beginning about 75 years ago (Lee 1985; Kolbe et al. 2004), and *A. cristatellus*, which was introduced from Puerto Rico about 40 years ago (Kolbe et al. 2012). We also tested for effects of local environment by comparing these two southern Florida populations with allopatric conspecifics in northern Florida (*A. sagrei*) and Puerto Rico (*A. cristatellus*). Finally, to assess possible sources of intraspecific variation in microbiota, we conducted two further studies on *A. sagrei*. First, we tested for temporal variation in bacterial

communities of free-living individuals by sampling, releasing, and resampling the same animals within a week of initial capture. Second, we compared the microbiota of free-living *A. sagrei* to those of individuals that we maintained in captivity for over a year on a simplified diet of domestic crickets. This allowed us to assess both overall dietary effects on the microbiota and the extent of variation among individuals when controlling for any differences in diet.

Results

General patterns in *Anolis* microbiota

We analyzed bacterial composition in 121 anole fecal samples (Table 1). From these samples, we generated a total of 2,662,283 high-quality reads, yielding a median of 15,157 reads per sample. Bacterial communities of anoles were complex, averaging 105 unique 97% OTUs per 1,000 sequences. Species richness estimated by Chao1 varied substantially among samples from 22 to 1,106 (mean = 209, SD = 137), and the Shannon index also varied substantially from 0.14 to 6.57 (mean = 3.63, SD = 1.60). We also observed considerable variation in bacterial composition, such that only 8 OTUs were shared by greater than 50% of wild anole samples (3 *Bacteroides* spp., 1 *Citrobacter* sp., *Clostridium perfringens, Eubacterium dolichum,* and 2 unclassified taxa in Peptostreptococcaceae and Lachnospiraceae). The average Jaccard distance between pairs of samples collected in the wild was 0.93, which means that, on average, any two samples only shared 7% of their bacterial OTUs. Exclusion of the rarest 5% of OTUs lowered the average Jaccard distance to 0.91. Taxonomic assignment revealed representatives from 22 bacterial phyla and 251 genera (Fig. 2; Appendix S2). The vast majority of sequences (95%) belonged to the bacterial phyla Firmicutes (61.1% of reads), Proteobacteria (19.1%), and Bacteroidetes (14.8%). Of the 251 genera identified, the most abundant were: *Bacteroides*

(10.1%), *Citrobacter* (9.6%), *Clostridium* (8.8%), *Lactococcus* (4.6%), *Parabacteroides* (3.0%), *Eubacterium* (2.5%), *Enterococcus* (2.3%), *Bacillus* (2.2%), *Dorea* (2.1%), *Blautia* (2.0%), *Staphylococcus* (1.7%) and *Enterobacter* (1.6%).

Comparison of ecomorphs on Puerto Rico

Microbiota were highly variable both within and among Puerto Rican species and ecomorphs when compared at the level of bacterial phylum, family, and genus (Fig. 2A). We found no differences in alpha diversity among ecomorphs or species (nested within ecomorph) when assessed using either the Chao1 index (ecomorph: $F_{2,75} = 0.64$; $P = 0.53$; species: $F_{3,75} = 0.26$, $P = 0.85$; Fig. 3A) or the Shannon index (ecomorph: $F_{2,75} = 0.15$, $P = 0.86$; species: $F_{3,75} = 0.26$; $P = 0.85$; Fig. 3B). Moreover, visual inspection of PCoA plots revealed no obvious clustering of bacterial beta diversity by ecomorph or species (Fig. 3C-D). When comparing within- and between-group Jaccard and UniFrac distances, we found slightly but significantly lower distances within species than between species (Fig. 3E-F), though distances within ecomorphs were equivalent to distances between ecomorphs (after removing all within-species comparisons; Fig. S3). These results were consistent irrespective of whether we included or excluded two species (*A. gundlachi* and *A. pulchellus*; Fig. S3) exhibiting high within-species distances that also gave rise to high between-species distances (Fig. S4). Accounting for OTU abundance using weighted UniFrac distances tended to homogenize within-species distances across species and reduce the difference in within- versus between-species distances (Fig. 3F), indicating that rare OTUs were driving differences in microbiota among species. Removal of the rarest 5% of bacterial OTUs had a similar effect in reducing the difference in within- versus between-species distances (Fig. S5), but all other patterns in beta diversity remained essentially unchanged when excluding rare OTUs (Fig. S5) Mantel tests revealed a weak but significant

association between genetic distances calculated from the *Anolis* phylogeny and both Jaccard (r = 0.1299; $P < 0.001$; excluding rare OTUs: r = 0.1291; $P < 0.001$) and unweighted Unifrac distances (r = 0.1135; $P < 0.001$; excluding rare OTUs: r = 0.1068; $P < 0.001$), indicating a weak effect of host phylogeny on composition of the microbiota.

Comparison of convergent ecomorphs in sympatry and allopatry

Comparison of two distantly related trunk-ground ecomorphs (*A. cristatellus* and *A. sagrei*) in sympatry and allopatry revealed pronounced effects of species and location on alpha and beta diversity of their bacterial communities. In southern Florida, where both species are sympatric, *A. cristatellus* exhibited a higher Chao1 index ($F_{1,20}$ = 6.92; P = 0.016; Fig. 4B), but the two species did not differ in Shannon indices ($F_{1,20}$ = 0.14; P = 0.710; Fig. 4C). PCoA plots for these sympatric populations revealed distinct clustering of beta diversity by species (Fig. 4D). Within-species distances for *A. cristatellus* were significantly lower than those for *A. sagrei* when measured as Jaccard distances (Fig. 4G) and as unweighted Unifrac distances ($P < 0.0001$; data not shown), but not as weighted Unifrac distances (Fig. 4G), indicating that species differences were driven in part less abundant bacterial OTUs. Nonetheless, these patterns persisted even after exclusion of the rarest 5% of OTUs. Comparison of within- and between-species distances revealed that *A. cristatellus* individuals were more similar to one another than to heterospecific *A. sagrei* individuals (Fig. 4G).

In allopatric comparisons, the population of *A. cristatellus* from southern Florida exhibited a higher Chao1 index ($F_{1,27}$ = 10.97; P = 0.002; Fig. 4B) and Shannon index ($F_{1,27}$ = 13.14; P = 0.001; Fig. 4C) than the population from Puerto Rico. Likewise, the population of *A. sagrei* from southern Florida exhibited a marginally higher Chao1 index ($F_{1,22}$ = 4.33; P = 0.049; Fig. 4B) and Shannon index ($F_{1,22}$ = 3.44, P = 0.077; Fig. 4C) than the population from northern

Florida, though these minor population differences in *A. sagrei* were not significant following correction for multiple comparisons (Fig. 4B-C). PCoA plots revealed pronounced clustering by geographic location in *A. cristatellus* (Fig. 4F), but not in *A. sagrei* (Fig. 4E). Regardless of the measure of beta diversity, *A. sagrei* individuals from northern Florida were more similar to one another than to individuals from southern Florida, whereas *A. sagrei* individuals from southern Florida were no more similar to one another than to conspecifics from northern Florida (Fig. 4H). Likewise, *A. cristatellus* individuals from southern Florida were more similar to one another than to individuals from Puerto Rico, whereas individuals from Puerto Rico were no more similar to one another than to conspecifics from southern Florida when using Jaccard distances (Fig. 4I). Only when using weighted UniFrac distances in *A. cristatellus* (Fig. 4I) did we observe a consistent tendency for conspecifics within each population to resemble one another more strongly than they resembled conspecifics from another population. Patterns in beta diversity for allopatric comparisons remained essentially unchanged when excluding the rarest 5% of OTUs.

Temporal variation in individual gut microbiota

Resampling of five *A. sagrei* individuals from northern Florida within a week of their initial capture revealed pronounced intra-individual variation in bacterial composition at the phylum and genus levels (Fig. S6). Analyses of beta diversity revealed that within-individual differences in bacterial OTUs were consistently lower than between-individual differences for Jaccard (mean within = 0.59, mean between = 0.88), unweighted UniFrac (0.46, 0.65), and weighted UniFrac distances (0.31, 0.48). Nonetheless, on average, only 72% of the same bacterial phyla and 53% of the same bacterial genera were present at both time points in any individual anole (average Jaccard distance = 0.28 at phylum level, 0.47 at genus level; Fig. S6).

Comparison of wild and captive anoles

Comparison of microbiota between wild *A. sagrei* and captive *A. sagrei* maintained in the laboratory on a simplified diet of domestic crickets revealed higher Shannon indices in wild anoles ($F_{1,23} = 8.82$, $P = 0.007$), but no difference in Chao1 indices ($F_{1,23} = 2.186$, $P = 0.153$). Individuals maintained in the laboratory also exhibited less variance in bacterial diversity relative to free-living *A. sagrei* (Brown-Forsythe test for unequal variances in Shannon indices: $F_{1,22} = 5.79$: $P = 0.025$). PCoA analysis revealed a clear separation of lab and free-living *A. sagrei* along PC1, though captive anoles still harbored microbiota that were distinct from their cricket diet (Fig. 5A). Within-group distances were lower in the lab than in free-living anoles, irrespective of whether they were calculated as Jaccard or UniFrac distances, and within-group distances were lower than between-group distances using Jaccard and unweighted UniFrac metrics (Fig. 5B).

Discussion

We examined 121 anoles representing 7 species from Puerto Rico and Florida to test whether and how the ecological and evolutionary diversification of a host lineage influences the composition of its bacterial communities. We found that populations and species differed subtly in both alpha and beta diversity of their microbiota, and that species differences in beta diversity were associated with genetic distances estimated from the host phylogeny, as observed in other taxa (Yildirim *et al.* 2010; Degnan *et al.* 2012; Sanders *et al.* 2014). Nonetheless, the major finding to emerge from our study is that bacterial diversity and intraspecific variation in community composition are strikingly high for the microbiota of *Anolis* lizards. On average, any two conspecific anoles from Puerto Rico shared only 10% of their bacterial OTUs, and less than 1% of OTUs appeared in greater than 50% of all individual samples. Moreover, differences in bacterial composition between conspecific individuals were generally comparable to those

between heterospecific individuals. In comparison, nestmates of turtle ants (genus *Cephalotes*) shared an average of 34% of their bacterial OTUs and this consistency in gut microbiota extended from the colony to the genus level of the host (Sanders et al. 2014). Accordingly, in turtle ants, there was a strong correlation between gut microbiota and the host phylogeny (Mantel test $r \approx 0.5$) (Sanders et al. 2014). In contrast, we found that the association between bacterial composition and host phylogeny was much weaker in a clade of *Anolis* lizards from Puerto Rico (Mantel test $r \approx 0.1$), presumably due to high bacterial variation within each species, rather than low variation across species (e.g., Fig 2; Appendix S2).

In other host species, changes in habitat, such as those induced by deforestation, can significantly alter gut microbiota on relatively short ecological timescales (Amato et al. 2013). Over longer evolutionary timescales, broad convergence in gut microbiota has been documented in association with the evolution of myrmecophagy across several mammalian lineages (Delsuc et al. 2014), and other broad dietary classifications (e.g., carnivory, herbivory, omnivory) also explain significant variation in mammalian gut microbiota (Ley et al. 2008; Muegge et al. 2011; Delsuc et al. 2014). Likewise, gut microbiota of fruit-feeding *Drosophila* species differ from those of flower-feeding species (Chandler et al. 2011). These examples of habitat and diet shaping gut microbiota on ecological and evolutionary timescales stand in contrast to our finding that microbiota of *Anolis* lizards did not differ in any obvious fashion with respect to host ecomorph. Moreover, sympatric members of the same ecomorph class often differed in bacterial beta diversity, despite their presumed similarity in habitat use (Fig. 3; Fig. 4D,G). Major divergences in the evolution of this *Anolis* clade on Puerto Rico likely date back millions to tens of millions of years (Brandley & de Queiroz 2004; Losos 2009). Therefore, the lack of obvious separation of *Anolis* microbiota by ecomorph is unlikely to be the result of an insufficient evolutionary timescale for divergence.

We suggest that the lack of clear separation of microbiota by ecomorph is most likely due to high intraspecific variation in bacterial gut communities, which may in turn be due to a tendency for *Anolis* species to function as dietary generalists, despite their spatial segregation with respect to habitat niches (Losos 2009). Detailed analyses of stomach contents from Puerto Rican anoles reveal that most species eat a wide variety of arthropod taxa and other food items (e.g., snails, seeds), often with no particular prey item consistently dominating the diet (Wolcott 1923; Lister 1981; Losos 2009). More recently, sequencing of arthropod 16S DNA from fecal samples, which is analogous to our approach for bacterial 16S DNA, was used to characterize the diet of *Anolis sagrei* (Kartzinel & Pringle 2015). This approach revealed a diverse diet containing at least 217 molecular OTUs from nine arthropod orders, but only three of these OTUs were frequent enough to occur in >50% of the individuals sampled, whereas 180 were found in <5% of the individuals sampled (Kartzinel & Pringle 2015). This is analogous to our observation that only 19 of 2722 (0.7%) bacterial OTUs occurred in >50% of the individuals we sampled on Puerto Rico, whereas 2092 (77%) OTUs were found in <5% of individual anoles. If representative of other *Anolis* species, this high degree of variation in diet (Kartzinel & Pringle 2015) may help explain why fecal microbiota appear so variable among individual anoles.

Our comparison of two convergent trunk-ground anoles (*A. cristatellus* and *A. sagrei*) that are recently sympatric (i.e., within about 40 generations, Kolbe et al. 2012) in southern Florida illustrates that even ecologically similar species that share the same environment can differ substantially in their microbiota. Despite the broad ecological convergence between these two species, their lineages likely diverged over ten million years ago (Brantley and de Queiroz 2004; Losos 2009). This suggests that, over longer evolutionary timescales, *Anolis* evolution could more strongly impact the diversification of their microbiota, potentially via genetic divergence in host digestive and immune physiology. However, we cannot exclude the

possibility that differences in contemporary microbiota between these two invasive species could simply reflect the persistence (over roughly 40-80 generations) of distinct bacterial communities from the different source islands (Puerto Rico and Cuba) on which these two species evolved and from which they invaded southern Florida. However, in the absence of a clear mechanism for vertical transmission of microbiota from mother to offspring, this seems unlikely. We also found that geographic location had a significant effect on microbiota when comparing conspecific populations of these same two species in allopatry (Fig. 4). This may be due to the availability of different food sources across different local environments, as well as geographic variation in numerous other environmental factors, such as biotic habitat, rainfall and microclimate.

Although we cannot directly assess the role of diet in structuring geographic variation in *Anolis* microbiota, we did observe considerable temporal variation when re-sampling the same free-living individuals within a few days of initial capture (Fig. S5). In similar fashion, pronounced temporal variation in the microbiota of individuals was also observed in fecal samples from wild baboons (Ren et al. 2015). In Burmese pythons, bacterial diversity and community composition in the gut changed rapidly and dramatically within hours to days of feeding, and these changes were primarily due to shifts in the abundance of endogenous gut bacteria, rather than the introduction of new bacteria from the rodent meal (Costello et al. 2010). This agrees with our observation that captive anoles maintained on a simplified diet of domestic crickets retained microbiota that were distinct from the bacterial composition of their prey. We also found that these same captive *A. sagrei* had higher alpha diversity and lower variance in alpha and beta diversity of their bacterial communities, relative to wild conspecifics. This is similar to differences observed between laboratory and wild populations of fruit flies (Chandler et al. 2011). Nonetheless, we cannot definitively attribute these differences between captive and

wild anoles to diet per se, as they could reflect numerous other differences between the laboratory and natural environments.

Our analysis of 121 individuals representing 7 *Anolis* species provides the most comprehensive study to date of fecal or gut bacterial diversity in any reptile lineage (Costello *et al.* 2010; Hong *et al.* 2011; Lankau *et al.* 2012; Colston *et al.* 2015; McLaughlin *et al.* 2015). *Anolis* microbiota were dominated by the phyla Firmicutes (61.1% of reads), Proteobacteria (19.1%), and Bacteroidetes (14.8%), which collectively accounted for 95% of the reads we detected (Fig 2A; Appendix S2). This is broadly similar to patterns in mammals (Ley et al. 2008), and with the exception of alligators (Keenan *et al.* 2013), the gut microbiota of other reptiles also appear to be consistently dominated by Firmicutes and Bacteriodetes (Colston et al. 2015; Wehrle 2013; Costello et al. 2010; Hong et al. 2011;Yuan *et al.* 2015), although Proteobacteria range from the dominant microbial taxon in some studies (McLaughlin et al. 2015; Martin et al. 2010) to minor components of the reptile gut in others (Costello et al. 2010; Hong et al. 2011; Yuan *et al.* 2015). It will be illuminating to see whether future studies of reptiles and other ectothermic vertebrates reveal comparably high levels of intraspecific variation in gut microbiota. Despite this variation, we found a significant correlation between bacterial diversity and *Anolis* phylogeny, suggesting that the recent adaptive radiation of this host lineage on Puerto Rico has weakly influenced the diversification of their microbiota. Nonetheless, the variation that we observed within species was comparable to that observed between species, and we found no tendency for the microbiota to vary predictably as a function of ecomorph, potentially because most *Anolis* species function as dietary generalists.

Material and Methods

Sample collection

Dates, locations and sample sizes for individuals of each species are reported in Table 1. Only samples from adult male anoles were included in our study. We captured wild anoles by hand or noose and immediately placed each individual into an unused plastic sandwich bag, where it was held overnight at ambient temperature. The following day, we transferred a fecal pellet from the bag of each individual into a microcentrifuge tube, immediately froze each pellet at −20°C, and kept samples on ice (during transportation) or at −20°C until DNA extraction. Because we did not collect fecal pellets immediately upon defecation, it is possible that bacterial composition may have changed prior to preservation, and that variation among samples (but not among species or populations, which were sampled in identical fashion) may reflect unmeasured variation in the time between defecation and preservation. Bacterial communities sampled from feces are often qualitatively similar to those sampled directly from the gut or cloaca, though they may differ quantitatively (Colston et al. 2015; Stanley *et al.* 2015), so we use fecal samples as a proxy for gut microbiota while acknowledging these caveats.

For *A. sagrei* individuals that we recaptured to assess individual repeatability of the microbiota, we used toe clips to identify each individual between captures. For *A. sagrei* individuals that we maintained on a controlled diet in captivity, we initially collected wild adults from Great Exuma, Bahamas (23°29'N, 75°45'W) and transported them to the University of Virginia. We maintained these anoles individually in small, plastic terraria (40 x 23 x 32 cm; Lee's Kritter Keeper, San Marcos, CA) containing a potted plant, carpet substrate, and PVC tube for perching and hiding. We maintained constant 29°C diurnal temperature and 65% relative humidity, 13L:11D (breeding season) or 12L:12D (non-breeding season) photoperiod, and placed each cage under two ReptiSun 10.0 UVB bulbs (ZooMed, San Luis Obispo, CA) that

were illuminated during the light phase. Three times per week, we offered captive anoles 5-7 crickets (*Gryllodes sigillatus*; Ghann's Cricket Farm, Augusta, GA), dusted weekly with Fluker's Reptile Vitamin and Calcium supplements (Fluker's Cricket Farms, Port Allen, LA) and maintained on a diet of carrots, kale, sweet potatoes, and apples. We maintained these wild-caught anoles in captivity on this diet for over a year before collecting fecal samples. To assess the bacterial communities associated with crickets as a food source, we froze and homogenized 15 whole crickets, then extracted DNA from this homogenate.

DNA extraction and 16S rRNA gene sequencing

We extracted DNA in a 96-well format using ZR-96 Fecal DNA Kits (Zymo Research, Orange, CA) following the manufacturer's protocol, then amplified the V1-V3 hypervariable regions of the 16S rRNA gene using two primers containing the universal sequences 27F (5'-AGRGTTTGATCMTGGCTCAG-3') and 534R (5'-TTACCGCGGCTGCTGGCAC-3'). We added a unique 8bp barcode to each primer to tag the samples and used a 50 uL reaction for each PCR amplification by QIAGEN Taq polymerase (Qiagen Inc, CA). PCR conditions consisted of 94°C for 3 min, followed by 25 cycles of 94°C for 30 s, 57°C for 30 s, and 72°C for 60 s, with a final extension of 5 min at 72°C. We quantified 16S rRNA amplicons from different samples, pooled them in equal molar ratios, then gel purified and sequenced them on an Illumina MiSeq platform using the 300bp paired-end (PE) protocol. We performed all liquid transfer steps on a Biomek NXp liquid handling station (Beckman-Coulter Inc., Fullerton, USA).

Sequence processing, quality control, and OTU classification

We filtered the resulting sequences according to base quality using TRIMMOMATIC 0.32 with settings of LEADING = 3, SLIDINGWINDOW = 10:20, and MINLEN = 50 (Bolger

et al. 2014). Paired-end reads passing the quality filter were merged using FLASH (-r 301 -f 447 -s 45 -x 0.05)(Magoč & Salzberg 2011). The successfully merged reads were assigned to samples by barcodes and processed using the QIIME pipeline (Caporaso et al. 2010). We identified chimeric sequences using usearch (Edgar et al. 2011) implemented in QIIME with both de novo and reference-based detection algorithms, retaining only those sequences that were flagged as non-chimeras with both detection methods. We removed non-16S rRNA sequences using hmmsearch (Eddy 1998) against a custom-made 16S rRNA gene model and removed mitochondrial and chloroplast sequences using the Ribosomal Database Project (RDP) classifier (Wang et al. 2007). We clustered the remaining reads to operational taxonomic units (OTUs) using the centroid-based UCLUST algorithm (Edgar 2010) with a 97% identity threshold, then selected the most abundant sequence of each OTU as the representative sequence using the RDP classifier.

Analysis of microbiota

To remove heterogeneity due to sequencing effort, we rarefied samples to 3,000 reads. This resulted in the removal of 15 samples and produced a final dataset of 121 samples (Table 1). Rarefaction curves for each species or population are presented in Figure S1. Good's coverage estimates averaged 0.97 ± 0.02 (range 0.84-0.99) for all samples (species means ranged from 0.96-0.98) following rarefaction to 3,000 reads, indicating that the majority of the bacterial community was captured at this level of rarefaction. We rarefied the Florida subset (*A. cristatellus* and *A. sagrei* from northern and southern Florida) to 1,000 reads to increase the number of samples in the analysis, which added 7 samples that were excluded from the 3,000-read dataset (mean Good's coverage estimate = 0.94 ± 0.03). A Mantel test correlating distance matrices for the 1,000 and 3,000-read datasets revealed a high level of congruence (Jaccard

distance: $r = 0.90$, $P = 0.001$; unweighted UniFrac: $r = 0.85$, $P = 0.001$). We calculated descriptive and comparative statistics for microbiota using QIIME 1.9.1, unless otherwise specified.

We calculated two measures of alpha diversity (Chao1 and Shannon indices) for each sample using rarefied OTU tables. The Chao1 index estimates species richness based on the number of observed OTUs in each sample using the formula:

$$S_{chao1} = S_{obs} + \frac{n_1(n_1 - 1)}{2(n_2 + 1)}$$

where S_{chao1} is the estimated richness, S_{obs} is the observed number of species, n_1 is the number of singleton taxa (taxa represented by a single read in that community), and n_2 is the number of doubleton taxa. If a sample contains many singletons, it is likely that more undetected OTUs exist, and the Chao 1 index will estimate greater species richness than it would for a sample without rare OTUs. The Shannon index combines species richness (number of OTUs) and evenness (relative abundance of different OTUs) to produce a summary measure of species diversity. We tested for differences in Chao1 and Shannon indices among Puerto Rican anoles using two-way ANOVA with the measure of alpha diversity as the dependent variable and ecomorph and species (nested within ecomorph) as independent variables. Depending on the distribution of data, we used both ANOVA and non-parametric Kruskal-Wallis tests with post hoc comparisons to assess differences in alpha diversity among sympatric and allopatric populations of *A. cristatellus* and *A. sagrei*. We used *t*-tests to assess differences in mean alpha diversity and Brown-Forsythe tests to compare variance in alpha diversity between *A. sagrei* maintained in captivity on a controlled diet versus those from wild populations.

To compare beta diversity among samples, we first excluded any OTUs that were only represented by a single sequence read. We then performed Principal Coordinate Analysis (PCoA) on Jaccard, unweighted UniFrac, and weighted UniFrac distances (Lozupone & Knight

2005) calculated using 97%-similarity OTUs. Jaccard distance measures the dissimilarity of two communities based on bivariate classifications of the presence or absence of microbial OTUs. UniFrac distance takes phylogenetic relationships among OTUs into account, and weighted UniFrac distance further considers the relative abundances of OTUs. For each of these metrics, we calculated pairwise distances between all individual samples and then classified each pairwise distance as occurring within a given group (species, ecomorph, population) or between two groups. We then (1) tested for differences in within-group distances among different groups, (2) tested for differences in between-group distances among different pairwise combinations, and (3) compared within- and between-group distance to one another using non-parametric Mann-Whitney tests or Kruskal-Wallis tests with post hoc Dunn's multiple comparisons tests. To test whether the host phylogeny explained variation in microbiota, we used Mantel tests implemented in QIIME to assess congruence between bacterial community dissimilarities and host genetic distances, which we calculated from branch lengths in the *Anolis* phylogeny of Rabosky and Glor (2010). Rare OTUs can confound detection of meaningful patterns in bacterial diversity (e.g., Colston et al. 2015), so in addition to using measures of community dissimilarity weighted by OTU abundance (e.g., weighted UniFrac distance), we repeated the analyses described above with a dataset in which we excluded the rarest 5% of OTUs from each sample.

Author contributions

All authors designed the research, analyzed the data, and wrote the manuscript. TR and MW performed sequencing and bioinformatic analyses. AFK and RMC collected samples.

Acknowledgments

We thank M. Johnson, B. Kircher, M. Oberndorf, J. Strecula, J.Murray, A. Hanninen, E.D. Brodie III, and V. Formica for assistance collecting anoles. We thank J. Price, H. Donald-Cannon, and D. Rose for care of captive anoles. Samples were collected under permits from Estado Libre Asociado de Puerto Rico, Bahamas Ministry of Agriculture, Bahamas Environment Science and Technology Commission, Miami-Dade County Parks and Recreation, and Guana Tolomato Matanzas National Estuarine Research Reserve. All research was conducted in compliance with the University of Virginia Animal Care and Use Committee (protocol 3896). Funding was provided by the University of Virginia (startup funding to MW and RMC), and a Theodore Roosevelt Memorial Grant from the American Museum of Natural History (to AFK).

References

Amato KR (2013) Co-evolution in context: The importance of studying gut microbiomes in wild animals. *Microbiome Science and Medicine*, 10–29.

Amato KR, Yeoman CJ, Kent A *et al.* (2013) Habitat degradation impacts black howler monkey (Alouatta pigra) gastrointestinal microbiomes. *The ISME Journal*, **7**, 1344–1353.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Brandley MC, de Queiroz K (2004) Phylogeny, ecomorphological evolution, and historical biogeography of the *Anolis cristatellus* series. *Herpetological Monographs*, **18**, 90–126.

Caporaso JG, Kuczynski J, Stombaugh JI *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Chandler JA, Lang JM, Bhatnagar S, Eisen JA, Kopp A (2011) Bacterial communities of diverse *Drosophila* species: Ecological context of a host–microbe model system. *PLoS Genetics*, **7**,

e1002272.

Colston TJ, Noonan BP, Jackson CR (2015) Phylogenetic analysis of bacterial communities in different regions of the gastrointestinal tract of *Agkistrodon piscivorus*, the cottonmouth snake. *PLoS ONE*, **10**, e0128793.

Costello EK, Gordon JI, Secor SM, Knight R (2010) Postprandial remodeling of the gut microbiota in Burmese pythons. *The ISME Journal*, **4**, 1375–1385.

Degnan PH, Pusey AE, Lonsdorf EV *et al.* (2012) Factors associated with the diversification of the gut microbial communities within chimpanzees from Gombe National Park. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 13034–13039.

Delsuc F, Metcalf JL, Wegener Parfrey L *et al.* (2014) Convergence of gut microbiomes in myrmecophagous mammals. *Molecular Ecology*, **23**, 1301–1317.

Eddy SR (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.

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

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Falush D, Wirth T, Linz B *et al.* (2003) Traces of human migrations in *Helicobacter pylori* populations. *Science*, **299**, 1582–1585.

Hong P-Y, Wheeler E, Cann IKO, Mackie RI (2011) Phylogenetic analysis of the fecal microbial community in herbivorous land and marine iguanas of the Galápagos Islands using 16S rRNA-based pyrosequencing. *The ISME Journal*, **5**, 1461–1470.

Hooper LV, Littman DR, Macpherson AJ (2012) Interactions between the microbiota and the immune system. *Science*, **336**, 1268–1273.

Kartzinel TR, Pringle RM (2015) Molecular detection of invertebrate prey in vertebrate diets:

trophic ecology of Caribbean island lizards. *Molecular Ecology Resources*, **15**, 903–914.

Keenan SW, Engel AS, Elsey RM (2013) The alligator gut microbiome and implications for archosaur symbioses. *Scientific Reports*, **3**, 2877.

Kolbe JJ, Glor RE, Rodríguez Schettino L *et al.* (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.

Kolbe JJ, Vanmiddlesworth PS, Losin N, Dappen N, Losos JB (2012) Climatic niche shift predicts thermal trait response in one but not both introductions of the Puerto Rican lizard *Anolis cristatellus* to Miami, Florida, USA. *Ecology and evolution*, **2**, 1503–1516.

Lankau EW, Hong P-Y, Mackie RI (2012) Ecological drift and local exposures drive enteric bacterial community differences within species of Galápagos iguanas. *Molecular Ecology*, **21**, 1779–1788.

Lee JC (1985) *Anolis sagrei* in Florida: Phenetics of a colonizing species I. Meristic characters. *Copeia*, **1985**, 182-194.

Ley RE, Hamady M, Lozupone CA *et al.* (2008) Evolution of mammals and their gut microbes. *Science*, **320**, 1647–1651.

Lister BC (1981) Seasonal niche relationships of rain forest anoles. *Ecology*, **62**, 1548.

Losos JB (2009) *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles.* University of California Press, Berkeley CA.

Losos JB, Jackman TR, Larson A, de Queiroz K, Rodríguez-Schettino L (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, **279**, 2115–2118.

Lozupone CA, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, **71**, 8228–8235.

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

McLaughlin RW, Cochran PA, Dowd SE (2015) Metagenomic analysis of the gut microbiota of the Timber Rattlesnake, Crotalus horridus. *Molecular Biology Reports*, **42**, 1187–1195.

Muegge BD, Kuczynski J, Knights D *et al.* (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science*, **332**, 970–974.

Qin J, Li R, Raes J *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464**, 59–65.

Rabosky DL, Glor RE (2010) Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 22178–22183.

Ren T, Grieneisen LE, Alberts SC, Archie EA, Wu M (2015) Development, diet and dynamism: longitudinal and cross-sectional predictors of gut microbial communities in wild baboons. *Environmental Microbiology*, **18**, 1312–1325.

Sanders JG, Powell S, Kronauer DJC *et al.* (2014) Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes. *Molecular Ecology*, **23**, 1268–1283.

Schoener TW (1967) The ecological significance of sexual dimorphism in size in the lizard *Anolis conspersus*. *Science*, **155**, 474–477.

Schoener TW (1968) The *Anolis* lizards of Bimini: Resource partitioning in a complex fauna. *Ecology*, **49**, 704-726.

Stamps JA, Losos JB, Andrews RM (1997) A comparative study of population density and sexual size dimorphism in lizards. *American Naturalist*. **149,** 64-90.

Stanley D, Geier MS, Chen H, Hughes RJ, Moore RJ (2015) Comparison of fecal and cecal microbiotas reveals qualitative similarities but quantitative differences. *BMC Microbiology*, **15**, 51.

Turnbaugh PJ, Ley RE, Mahowald MA *et al.* (2006) An obesity-associated gut microbiome with

increased capacity for energy harvest. *Nature*, **444**, 1027–1031.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.

Williams EE (1972) The origin of faunas. Evolution of lizard congeners in a complex island fauna: A trial analysis. In: *Evolutionary Biology* (eds Dobzhansky T, Hecht MK, Steere WC), pp. 47–89. Springer US, Boston, MA.

Williams EE (1983) Ecomorphs, faunas, island size, and diverse end points in island radiations of *Anolis*. In: *Lizard Ecology: Studies of a Model Organism* (eds Huey RB, Pianka ER, Schoener TW), pp. 326–370. Harvard University Press, Cambridge, MA.

Wolcott GN (1923) The food of Porto Rican lizards. *Journal of the Department of Agriculture of Porto Rico*, **VII,** 5–43.

Yildirim S, Yeoman CJ, Sipos M *et al.* (2010) Characterization of the fecal microbiome from non-human wild primates reveals species-specific microbial Ccommunities. *PLoS ONE*, **5**, e13963.

Yuan ML, Dean SH, Longo AV *et al.* (2015) Kinship, inbreeding and fine-scale spatial structure influence gut microbiota in a hindgut-fermenting tortoise. *Molecular Ecology*, **24**, 2521–2536.

Tables and Figures

Table 1. Collection dates and localities for wild anoles from Puerto Rico and Florida. Sample sizes are reported for the total number of fecal samples before and after rarefaction to exclude samples with low sequencing coverage (i.e. < 3000 reads per sample).

Location	Species	Number of samples	Number after rarefaction	Collection date	Comparison
Puerto Rico	*A. stratulus*	17	17	June 2014	ecomorph
Puerto Rico	*A. evermani*	15	15	June 2014	ecomorph
Puerto Rico	*A. krugi*	12	12	June 2014	ecomorph
Puerto Rico	*A. pulchellus*	10	9	June 2014	ecomorph
Puerto Rico	*A. gundlachi*	13	13	June 2014	ecomorph
Puerto Rico	*A. cristatellus*	16	15	June 2014	ecomorph, allopatry
Miami, FL (southern FL)	*A. cristatellus*	13	8	May 2014	sympatry, allopatry
Miami, FL (southern FL)	*A. sagrei*	11	8	May 2014	sympatry, allopatry, captive/wild
Palm Coast, FL (northern FL)	*A. sagrei*	17	12	July 2015	allopatry, repeatability, captive/wild
Lab	*A. sagrei*	12	12		captive/wild
Lab	Cricket	1	1		captive/wild

Figure 1. (A) Illustration of the six *Anolis* ecomorphs, depicting habitat partitioning and major morphological differences (modified from Williams 1983; Losos 2009). (B) Phylogenetic relationships and ecomorph classifications for the seven *Anolis* species sampled in this study, including the "outgroup" *A. sagrei* (native to Cuba and the Bahamas) and six species from a clade that evolved on Puerto Rico (phylogeny based on Rabosky & Glor 2010). Asterisks indicate the two trunk-ground ecomorphs that have invaded southern Florida from Cuba (*A. sagrei*) and Puerto Rico (*A. cristatellus*).

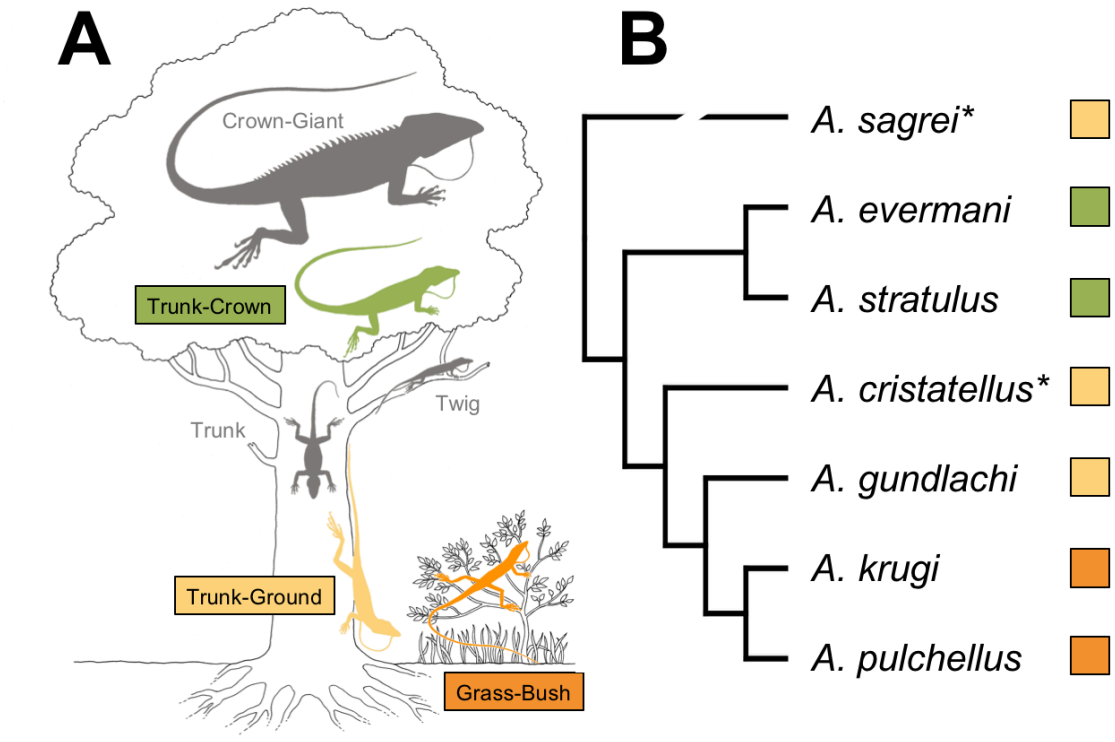


Figure 2. Diversity of *Anolis* gut microbiota as a function of host phylogeny. Each thin horizontal bar represents an individual lizard, with bacterial diversity (proportion of reads) coded at phylum, family, and genus (see Appendix S2 for key to microbial taxa).

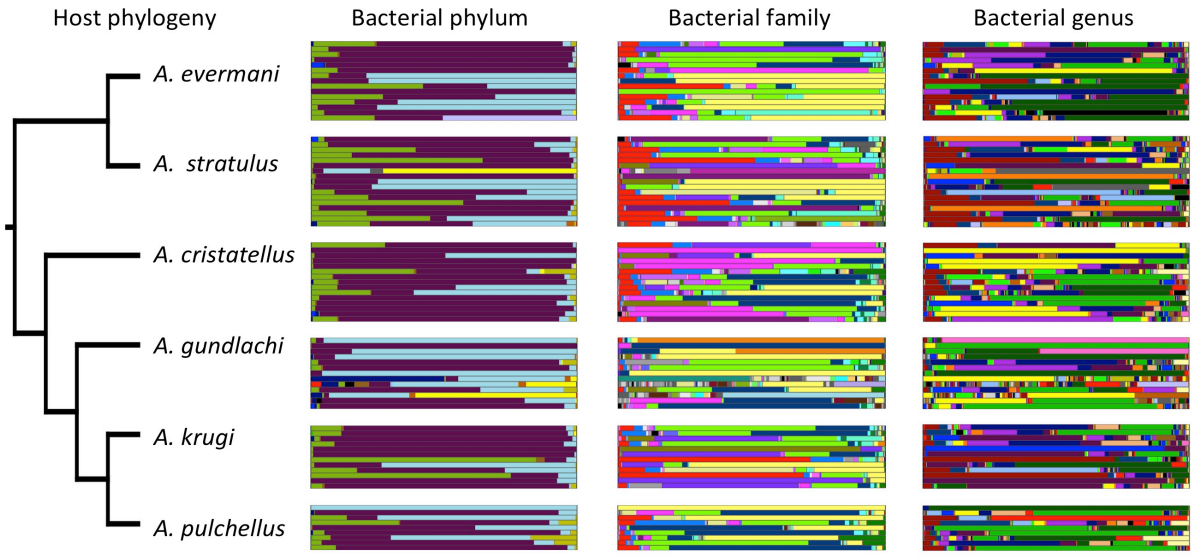


Figure 3. (A-B) Box-and-whisker plots illustrating medians, interquartiles, and ranges of bacterial alpha diversity expressed using (A) Chao1, and (B) Shannon indices for each host species, with bacterial OTUs assigned at 97% sequence similarity. (C-D) Distribution of gut microbiota across individual lizards, coded by species and ecomorph, as a function of the first three principle coordinate axes based on Jaccard distances. (E-F) Box-and-whisker plots (medians, interquartiles, and 10-90% percentiles) for within-species (E) Jaccard, and (F) weighted UniFrac distances (colored boxes) alongside comparisons of all within-species (W) to all between-species (B) distances (gray boxes). Letters denote statistical separation based on post hoc Dunn's multiple comparison tests. Asterisks indicate statistical significance at $P < 0.0001$ (***) and $P < 0.05$ (*).

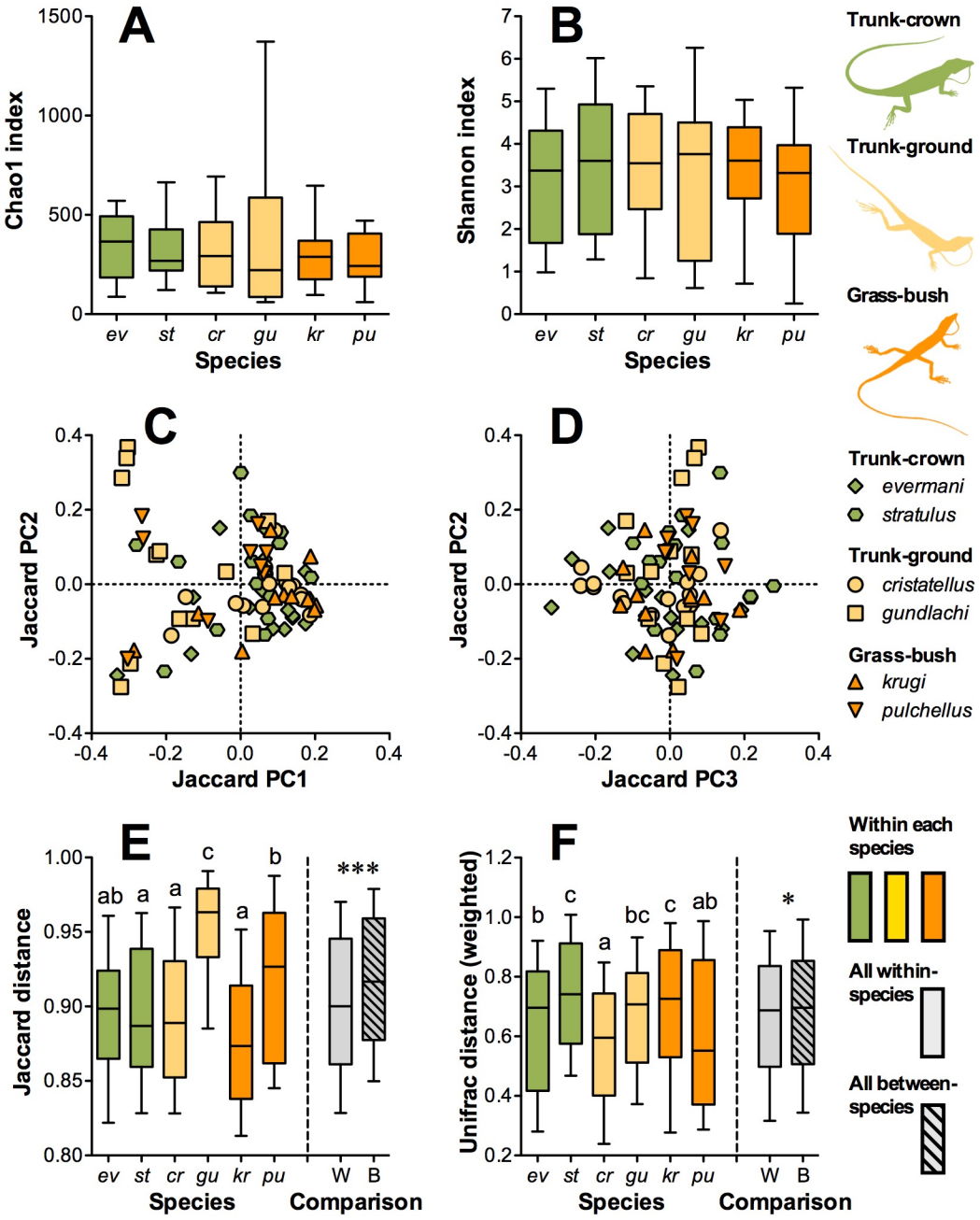

A
Chao1 index
1500
1000
500
0
ev st cr gu kr pu
Species
B
Shannon index
7
6
5
4
3
2
1
0
ev st cr gu kr pu
Species
Trunk-crown
Trunk-ground
Grass-bush
C
Jaccard PC2
0.4
0.2
0.0
-0.2
-0.4
-0.4 -0.2 0.0 0.2 0.4
Jaccard PC1
D
Jaccard PC2
Jaccard PC3
Trunk-crown
evermani
stratulus
Trunk-ground
cristatellus
gundlachi
Grass-bush
krugi
pulchellus
E
Jaccard distance
1.00
0.95
0.90
0.85
0.80
ab a a c a b
***
ev st cr gu kr pu W B
Species
Comparison
F
Unifrac distance (weighted)
1.2
1.0
0.8
0.6
0.4
0.2
b c a bc c ab
*
ev st cr gu kr pu W B
Species
Comparison
Within each species
All within-species
All between-species

Figure 4. (A) Locations of sympatric and allopatric populations of *A. cristatellus* and *A. sagrei*. (B-C) Box-and-whisker plots (medians, interquartiles, 10-90% percentiles) for two measures of alpha diversity within each population: (B) Chao1 index, and (C) Shannon index. (D-F) Principal coordinate analyses for (D) sympatric populations of both species, (E) allopatric populations of *A. sagrei*, and (F) allopatric populations of *A. cristatellus*. (G-H) Box-and-whisker plots (medians, interquartiles, 10-90% percentiles) for within- (solid boxes) and between-population (lined boxes) pairwise comparisons using Jaccard distance (left panels) and weighted UniFrac distance (right panels). Significance values are from non-parametric Kruskal-Wallis tests and lowercase letters indicate post hoc separation based on Dunn's multiple comparison tests.

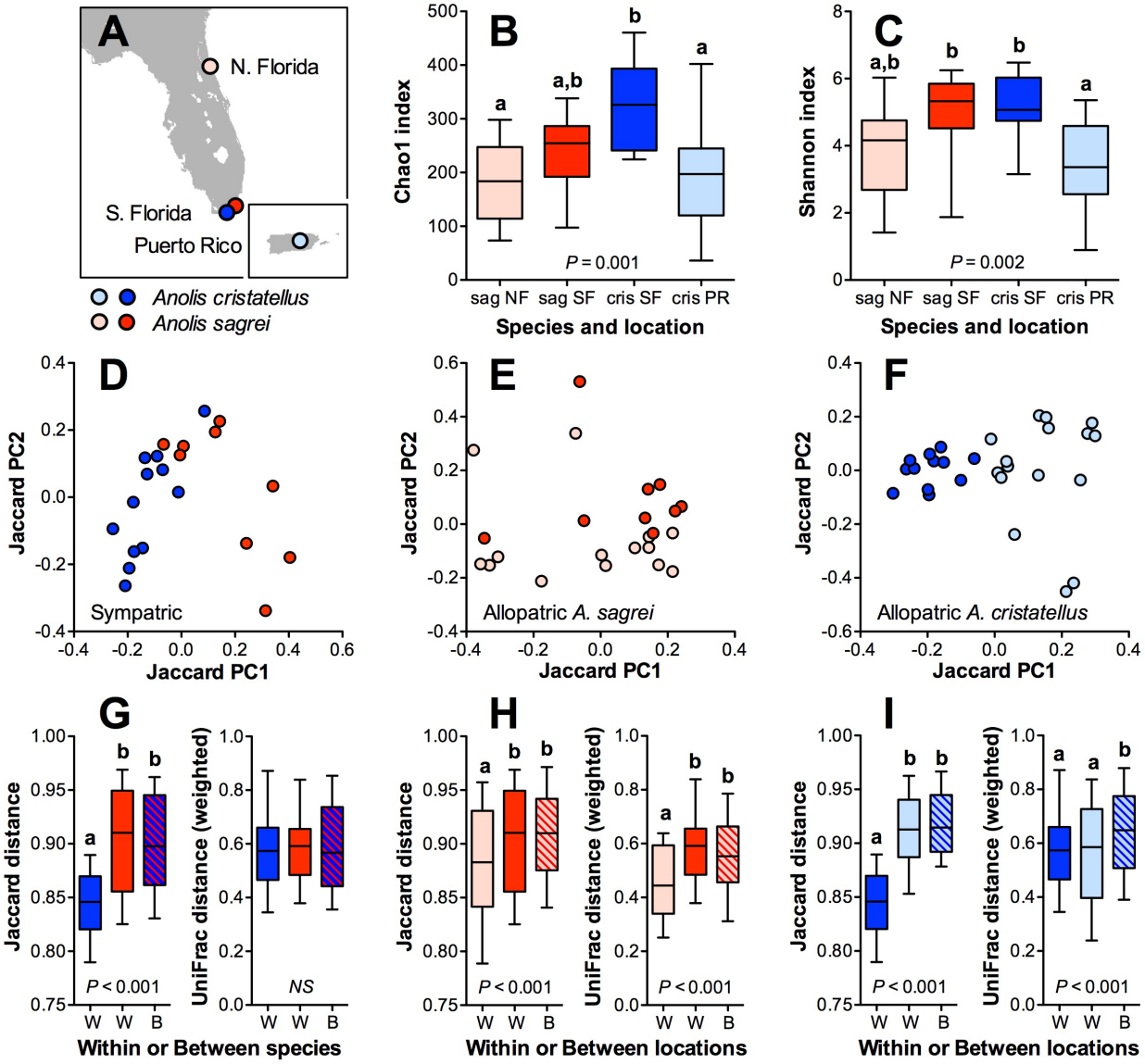
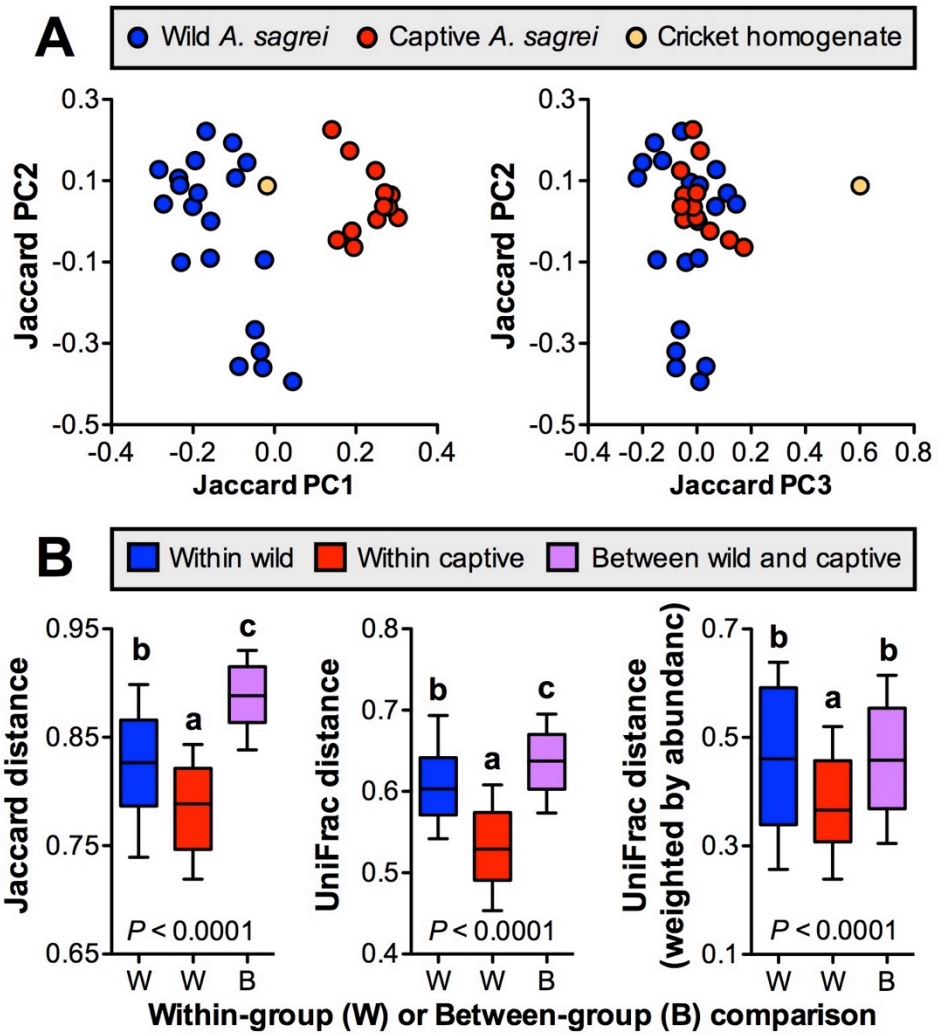
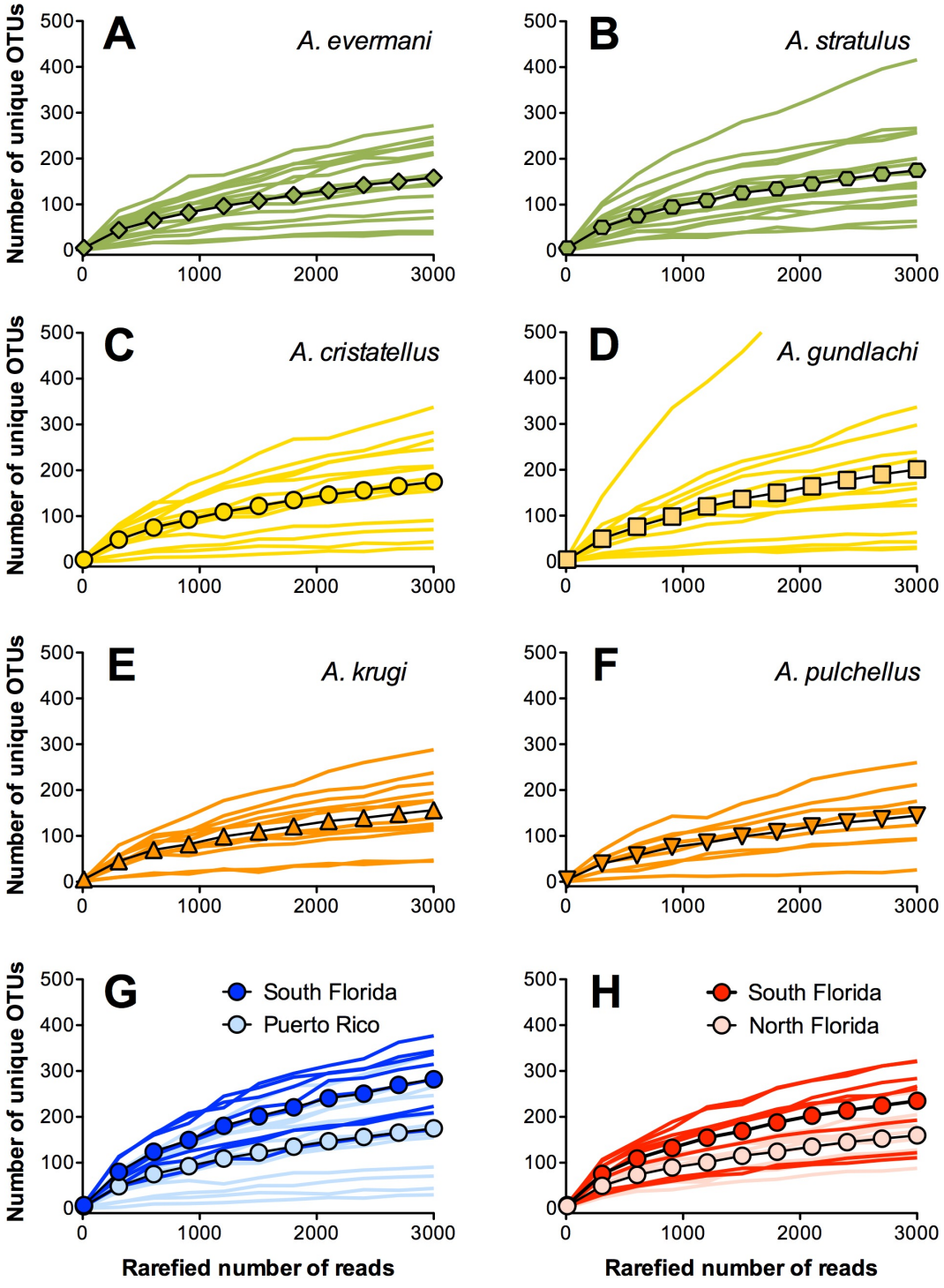

A
N. Florida
S. Florida
Puerto Rico
Anolis cristatellus
Anolis sagrei
B
Chao1 index
P = 0.001
sag NF sag SF cris SF cris PR
Species and location
C
Shannon index
P = 0.002
sag NF sag SF cris SF cris PR
Species and location
D
Jaccard PC2
Jaccard PC1
Sympatric
E
Jaccard PC2
Jaccard PC1
Allopatric A. sagrei
F
Jaccard PC2
Jaccard PC1
Allopatric A. cristatellus
G
Jaccard distance
P < 0.001
UniFrac distance (weighted)
NS
W W B
Within or Between species
H
Jaccard distance
P < 0.001
UniFrac distance (weighted)
P < 0.001
W W B
Within or Between locations
I
Jaccard distance
P < 0.001
UniFrac distance (weighted)
P < 0.001
W W B
Within or Between locations

Figure 5. (A) Principal coordinate analyses comparing gut microbiota of wild *A. sagrei* individuals to captive *A. sagrei* individuals maintained on a controlled diet of domestic crickets, alongside the bacterial community of a homogenate prepared from this food source. (B) Box-and-whisker plots (medians, interquartiles, 10-90% percentiles) for pairwise distances (Jaccard, unweighted UniFrac, and weighted UniFrac) calculated between individual lizards and binned into within- and between-group comparisons. Significance values are from Kruskal-Wallis tests and lowercase letters denote post hoc separation based on Dunn's multiple comparison tests.



Supplementary information

Figure S1. Rarefaction curves illustrating the increase in number of unique bacterial OTUs (97% sequence similarity) as a function of number of reads used for rarefaction. Each colored line is a sample from an individual and the symbols and black lines illustrate species or population means.



Appendix S2. Key to the colors for microbial taxa in Figs. 2A and S5. Numbers report frequency of reads for common taxa across samples from Puerto Rico. Numbers in parentheses report frequencies across the entire dataset (Puerto Rico and Florida).

A. Phylum

- 38.66% (60.75%) Firmicutes
- 35.23% (20.56%) Proteobacteria
- 19.80% (13.18%) Bacteroidetes
- 2.89% (0.13%) TM7
- 2.22% (1.25%) Actinobacteria
- 1.20% (4.13%) Other (Phyla with less than 1%)

B. Family

- 17.77% (15.71%) Enterobacteriaceae (phylum: Proteobacteria)
- 16.92% (18.60%) Lachnospiraceae (phylum: Firmicutes)
- 13.15% (11.01%) Peptostreptococcacea (phylum: Firmicutes)
- 10.37% (10.80%) Other (Families with <1% reads)
- 7.81% (9.81%) Clostridiaceae (phylum: Firmicutes)
- 7.58% (9.34%) Bacteroidaceae (phylum: Bacteroidetes)
- 6.74% (5.05%) Streptococcaceae (phylum: Firmicutes)
- 3.94% (3.09%) Enterococcaceae (phylum: Firmicutes)
- 2.94% (3.07%) Porphyromonadace (phylum: Bacteroidetes)
- 2.49% (1.86%) Staphylococcaceae (phylum: Firmicutes)
- 2.39% (2.49%) Bacillaceae (phylum: Firmicutes)
- 2.21% (3.87%) Erysipelotrichaceae (phylum: Firmicutes)
- 1.93% (1.62%) Coxiellaceae (phylum: Proteobacteria)
- 1.46% (1.58%) Unclassified family (phylum: Firmicutes)
- 1.25% (1.50%) Ruminococcaceae (phylum: Firmicutes)
- 1.05% (1.14%) Verrucomicrobiaceae (phylum: Verrucomicrobia)

C. Genus

- 19.98% (26.95%) Other (Genera with <1% reads)
- 13.24% (10.62%) *Citrobacteria* (phylum: Proteobacteria; family: Enterobacteria)
- 12.70% (10.64%) Unclassified Peptostreptococcacae (phylum: Firmicutes; family: Peptostreptococcacae)
- 7.77% (7.23%) Unclassified Lachnospiraceae (phylum: Firmicutes; family: Lachnospiraceae)
- 7.65% (9.41%) *Clostridium* (phylum: Firmicutes; family: Clostridiaceae)
- 7.58% (9.34%) *Bacteroides* (phylum: Bacteroidetes; family: Bacteroidaceae)
- 6.74% (5.05%) *Lactococcus* (phylum: Firmicutes; family: Staphylococcaceae)
- 4.40% (5.55%) Unclassified Lachnospiraceae (phylum: Firmicutes; family: Lachnospiraceae)
- 3.16% (2.51%) *Enterococcus* (phylum: Firmicutes; family: Enterococcaceae)
- 2.84% (2.96%) *Parabacteroides* (phylum: Bacteroidetes; family: Porphyromona)
- 2.49% (1.86%) *Staphylococcus* (phylum: Firmicutes; family: Staphylococcaceae)
- 2.38% (2.48%) *Bacillus* (phylum: Firmicutes; family: Bacillaceae)
- 1.93% (1.61%) *Rickettsiella* (phylum: Proteobacteria; family: Coxiellacae)
- 1.82% (1.81%) *Blautia* (phylum: Firmicutes; family: Lachnospiraceae)
- 1.71% (1.65%) *Enterobacter* (phylum: Proteobacteria; family: Enterobacteria)
- 1.68% (2.18%) *Dorea* (phylum: Firmicutes; family: Lachnospiraceae)
- 1.46% (1.58%) Unclassified Firmicutes (phylum: Firmicutes; unclassified family)
- 1.20% (0.97%) Unclassified Enterobacteria (phylum: Proteobacteria; family: Enterobacteria)
- 1.02% (1.13%) *Akkermansia* (phylum: Verrucomicrobia; family: Verrucomicrobiaceae)

Figure S3. Comparison of within-group to between-group distances (median, interquartiles, 10-90% percentiles) with groups defined as species (left) or ecomorphs (right), using Jaccard, unweighted UniFrac, and weighted (by abundance) UniFrac distances between individual anoles. Results are broadly similar whether including all species or excluding two species (*A. gundlachi* and *A. pulchellus*) characterized by high within- and between-species distances (see Fig. S3). Statistics are from Mann-Whitney U tests. Ecomorph comparisons exclude distances between individuals of the same species.

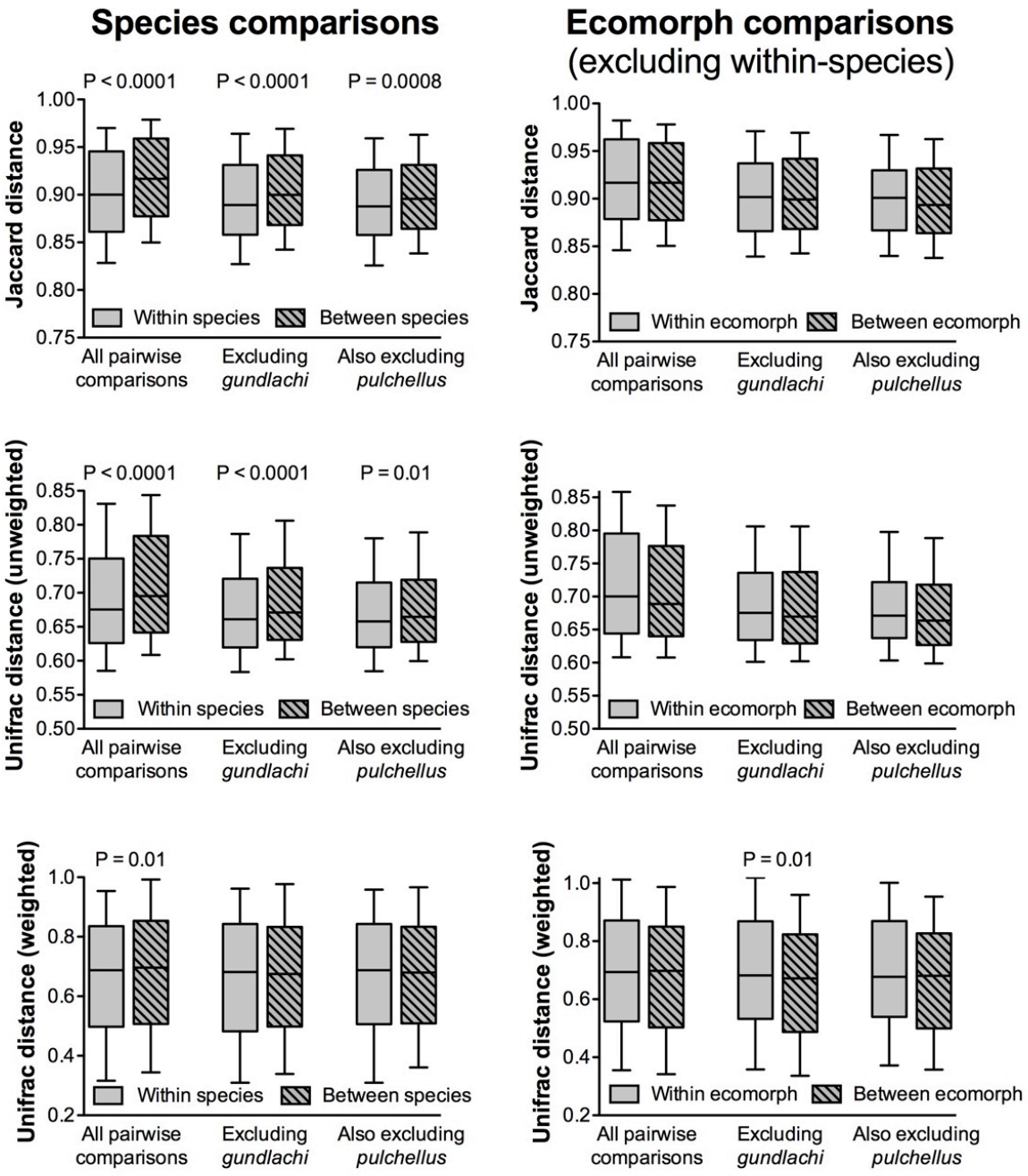


Figure S4. Box-and whisker plots (median, interquartile, 10-90% percentile) illustrating pairwise between-species measures of Jaccard, unweighted UniFrac, and weighted UniFrac distances. Colored boxes present within-species distances for comparison.

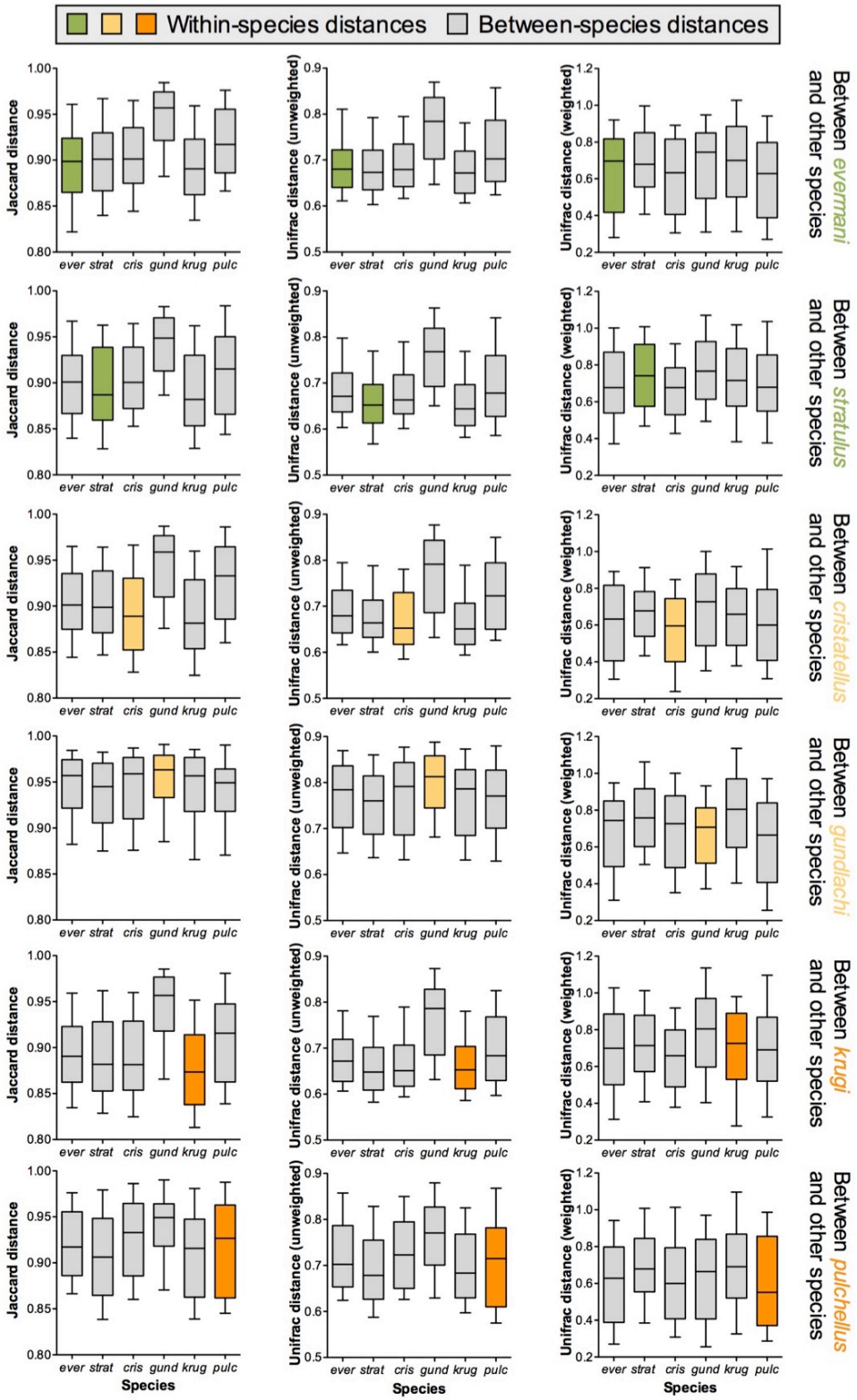


Figure S5. Comparison of results using all bacterial OTUs (full dataset reported in the main text) to those in which the rarest 5% of OTUs were removed from each sample (i.e., assigned a value of zero) prior to calculation of Jaccard distances. Exclusion of rare OTUs eliminated the trend for between-species Jaccard distances to exceed within-species Jaccard distances (compare A to B), similar to results obtained with UniFrac distances weighted by species abundance (compare to Fig. 3D). Otherwise, results were highly congruent between datasets and the removal of rare OTUs typically only decreased Jaccard distances (i.e., increased similarity between samples) by 1-2%. Likewise, PCoA plots, Mantel tests for phylogenetic similarity of bacterial communities, and comparisons of sympatric and allopatric populations of *A. sagrei* and *A. cristatellus* (Fig. 4) remained essentially unchanged when excluding rare OTUs (data not shown). Panel A is presented in the main text as Fig. 3E and panel C is presented in Fig. S3 – they are reproduced here for comparison between datasets.

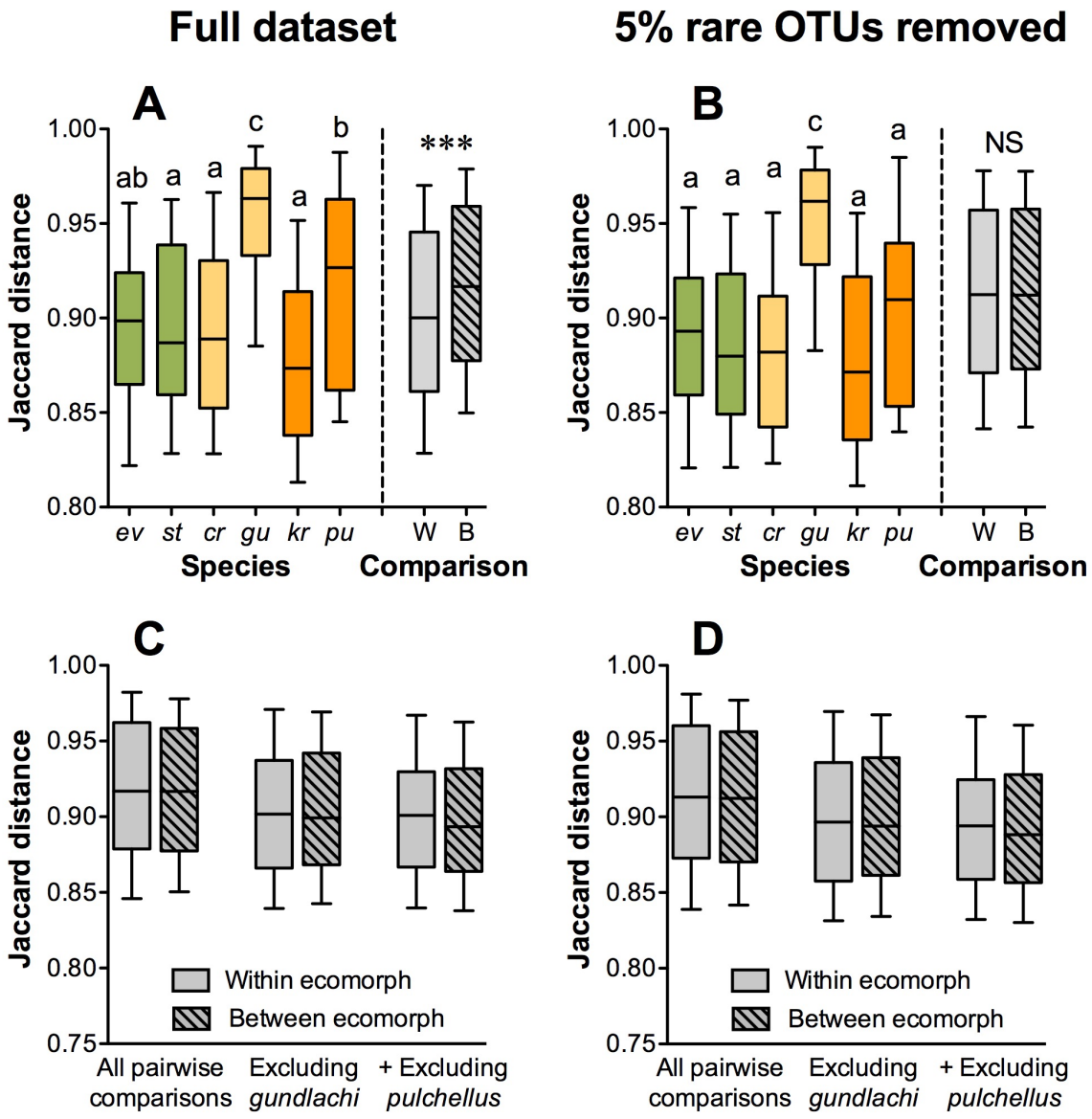
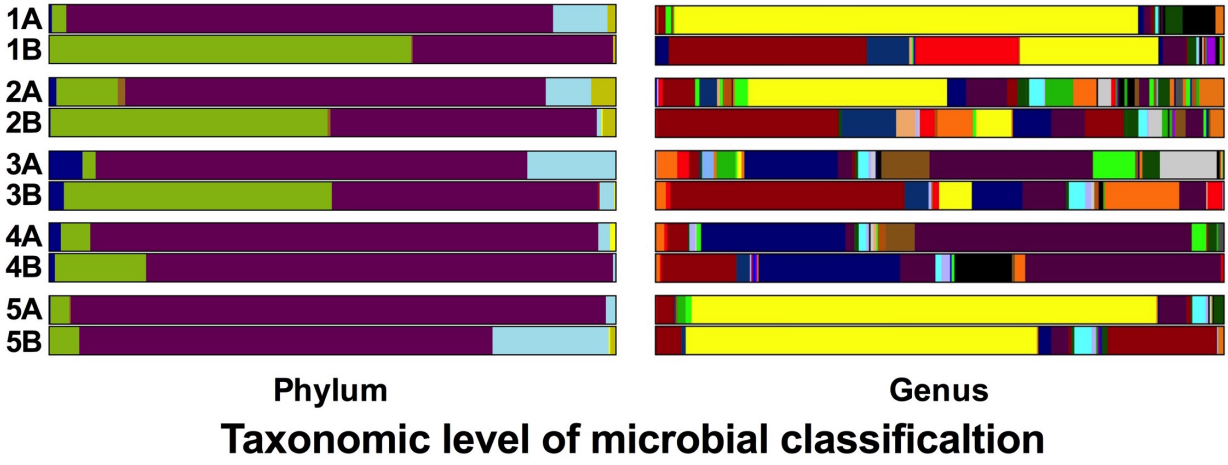


Figure S6. Representative examples of short-term changes in the gut microbiota of five free-living *Anolis sagrei* individuals (numbered 1-5) captured and sampled (A), then released, re-captured, and re-sampled within a week of initial capture (B), as assessed at the level of microbial phylum and genus. See Appendix S1 for key to microbial taxa.



Appendix 1. 16S rRNA survey revealed complex bacterial communities and evidence of bacterial interference on human adenoids

Formatted as a co-authored manuscript and published as:

Ren T, Glatt DU, Nguyen TN, Allen EK, Early SV, Sale M, Winther B, and Wu M. (2013) 16S rRNA survey revealed complex bacterial communities and evidence of bacterial interference on human adenoids. *Environmental Microbiology*, **15**(2): 535-547.

DOI: 10.1111/1462-2920.12000

Abstract

Adenoid microbiota plays an important role in the development of various infectious and noninfectious diseases of the upper airways, such as otitis media, adenotonsillitis, rhinosinusitis, and adenoid hypertrophy. Studies have suggested that adenoids could act as a potential reservoir of opportunistic pathogens. However, previous bacterial surveys of adenoids were mainly culture based and therefore might only provide an incomplete and potentially biased assessment of the microbial diversity. To develop an in-depth and comprehensive understanding of the adenoid microbial communities and test the "pathogen reservoir hypothesis", we carried out a 16S rRNA based, culture-independent survey of bacterial communities on 67 human adenoids removed by surgery. Our survey revealed highly diverse adenoid bacterial communities distinct from those of other body habitats. Despite large interpersonal variations, adenoid microbiota shared a core set of taxa and can be classified into at least five major types based on its bacterial species composition. Our results support the "pathogen reservoir hypothesis" as we found common pathogens of otitis media to be both prevalent and abundant. Co-occurrence analyses revealed evidence consistent with the bacterial interference theory in that multiple common pathogens showed "non-coexistence" relationships with non-pathogenic members of the commensal microflora.

Introduction

There is a growing recognition that microbes living on and inside our body (collectively called human microbiota) play important roles in human health and disease (Turnbaugh et al., 2007). For example, gut microbiota has been shown to aid in the metabolism of nutrient (Wostmann et al., 1983; Turnbaugh et al., 2006), outcompete the pathogenic bacteria and modulate the development of host immune system (Mazmanian et al., 2005). Recent culture-independent molecular surveys based on 16S rRNA pyrosequencing have revealed immense microbial diversity and detailed patterns of population variations that exist on our skin, inside our oral cavities and gastrointestinal and urogenital tracts (Eckburg et al., 2005; Costello et al., 2009; Grice et al., 2009; Ravel et al., 2011). These surveys showed that bacterial communities vary greatly between body sites and from individual to individual. Association studies have shown, however, that shifts in the microbial community structure can be associated with important human health conditions, including diabetes (Giongo et al., 2011), obesity (Ley et al., 2005; Ley et al., 2006; Cani et al., 2007), cancer (Turnbaugh et al., 2007), and cardiovascular disease (Ordovas and Mooser, 2006).

The nasopharyngeal tonsil, adenoid, is a lymphoid tissue located in the upper respiratory tract at the junction of nose and throat. The surface of the adenoids are extensive due to the folds and crypts (Winther and Innes, 1994), which are colonized by commensal microflora (Winther et al., 2009). As part of the immune system, adenoid plays a major role in body's immune response to infectious organisms that are introduced through the upper airways (Perry and Whyte, 1998). Macrophages and other white blood cells concentrate by the crypts in response to the microorganisms trapped there. Therefore, adenoid may serve the roles of both scout and sentry

for the immune system, by providing early exposure of immune system to pathogens. Because adenoid lies adjacently to the orifice of the Eustachian tube, the passage between the back of the nose and the inside of the ear, adenoid has long been recognized as an important factor in the pathogenesis of middle ear infection (also known as otitis media) by serving as a potential reservoir of opportunistic pathogens (Tomonaga et al., 1989; Faden et al., 1991; Bernstein et al., 1993; Faden et al., 1997; Bernstein, 1999; Dhooge et al., 1999; Brook et al., 2000; Karlidag et al., 2002; Marchisio et al., 2003; Nistico et al., 2011). Bacteria are postulated to spread via the short Eustachian tube to the middle ear where they cause acute, recurrent or chronic infections (Bluestone, 1999). Significant adenoid enlargement due to recurrent infections can also cause nasal obstruction, leading to breathing, swallowing, and sleep problems. If the condition does not improve with antibiotic therapy, surgical removal of adenoid is often recommended.

Despite its important roles in the etiology of otitis media, and in modulating the systematic and mucosal immunity, our knowledge of the adenoid microbiota is limited. Previous studies were either culture or PCR based surveys targeted to a few bacterial groups, and therefore only provided incomplete and potentially biased assessment of the microbial diversity (Brook et al., 2000; Fekete-Szabo et al., 2010; Khoramrooz et al., 2012). One recent study surveyed microbiota of middle ear, adenoid and tonsil of one individual using 16S rRNA pyrosequencing (Liu et al., 2011). It revealed a highly diverse adenoid microbial community that encompassed bacteria found in both the tonsil and the middle ear, supporting the hypothesis that adenoid may serve as a bacterial reservoir for both middle ear and tonsillar diseases. However, since only one patient was surveyed in that study, it is unclear whether conclusion from that study can be generalized.

It has been well documented that the indigenous bacterial flora of nasopharynx can inhibit the colonization of pathogens such as *Staphylococcus aureus*, *Streptococcus pneumonia* and *Haemophilus influenzae*, a phenomenon known as bacterial interference (Brook, 1999, 2005; Benninger et al., 2011). Culture-based quantitative studies have shown an inverse relationship between the abundance of viridans streptococci (non-pathogenic) and pathogens on adenoids (Bernstein, Sagahtaheri-Altaie et al. 1994). The inhibitory effect of viridans streptococci has also been demonstrated directly by *in vitro* assays when pathogens were grown in the presence of viridans streptococci or their culture filtrates (Bernstein et al., 1994; Tano et al., 1999; Bernstein et al., 2002; Tano et al., 2002; Bernstein et al., 2006).

In this study, we sought to develop an in-depth and comprehensive understanding of the adenoid microbial communities using a culture-independent molecular approach. Specifically, we employed 16S rRNA pyrosequencing to survey the diversity of adenoid microbiota, sampled from 67 individuals who underwent adenoidectomy for various reasons. We compared bacterial community structures between individuals to look for patterns that were common to adenoid microbiota and tested the pathogen reservoir and bacterial interference hypotheses.

Results

Bacterial species richness

To characterize the bacterial diversity present on the human adenoids, we surveyed the 16S rRNA V1-V2 hypervariable regions by multiplex pyrosequencing. We successfully PCR amplified and sequenced the 16S rRNA gene from 68 out of the 98 patients. One patient was removed from further analyses because of its relatively low sequence coverage. For patient 84,

the superior and inferior surfaces of adenoid were sampled independently and were treated as two separate samples. After filtering out the low quality reads and chimeric reads (see methods), the final data set contained 450,465 high-quality reads from 68 samples of 67 patients, with a median of 4,105 reads per sample.

To compare the bacterial species richness between samples, we carried out rarefaction analysis by plotting the number of species observed (as approximated using Operational Taxonomic Units (OTUs) at 97% identity cutoff) against the sequencing effort (Figure 1). For most samples, the curves start to level off, indicating that they have been well sampled with the sequencing effort. For some samples, the upward phase is still ongoing, indicating relatively higher species diversities in these samples. Consistent with this, Good's estimator shows relatively good coverage ranging from 80.8% to 99.3% (average 94.8%). This suggests that on average, five new species would be expected in every 100 additional 16S rRNA reads and that we have covered the vast majority of bacterial diversity on the 67 adenoids.

The species richness on adenoids was considerably high. In total, 3,121 distinct OTUs were observed. There were also extensive interpersonal disparities in bacterial richness. The number of observed bacterial species present in each patient varied substantially from 52 to 405 (median: 205). Accordingly, the overall bacterial diversity as measured by the Shannon index varied substantially from 0.79 to 6.60.

Most of the adenoid bacterial diversity existed below the species (97% OTU) level. This was revealed by plotting the number of OTUs against the identity cutoff values that were used to define OTUs (Supplementary Figure 1). The number of OTUs shows a dramatic flare-up around

the 99% identity cutoff. The presence of such a "hockey stick" figure has been previously observed in bacterial populations of the natural environment (Acinas et al., 2004; Brown and Fuhrman, 2005) and is also considered typical of human microbiota (Backhed et al., 2005; Bik et al., 2010).

Bacterial composition

The sequence reads were classified by the RDP classifier into 14 phyla (Figure 2) and 94 genera. The vast majority of sequences (96.8%) belonged to one of the nine phyla: Firmicutes, Proteobacteria, Fusobacteria, Actinobacteria, Bacteroidetes, Spirochaetes, Tenericutes, candidate division TM7 and SR1. Of the nine phyla, Firmicutes (45.4%), Proteobacteria (28.6%) and Fusobacteria (11.1%) were the most abundant. 3.2% of sequences were unclassified and therefore might represent novel lineages. Of the 94 genera identified, the most abundant genera were: Streptococcus (18.0%), Staphylococcus (14.7%), Haemophilus (11.2%), Fusobacterium (10.4%), Moraxella (5.7%), Prevotella (4.1%), Gemella (2.8%), Neisseria (2.7%), Corynebacterium (2.3%), Granulicatella (1.4%) and Pseudomonas (1.3%).

Similar to the findings of previous human microbiota studies (Eckburg et al., 2005; Costello et al., 2009; Grice et al., 2009; Yatsunenko et al., 2012), bacterial composition was highly variable between patients. To quantify the beta-diversity of the adenoid microbiota, we calculated the pairwise Chao-Jaccard abundance-based similarity index between samples. Chao-Jaccard similarity is based on the probability that two randomly chosen species, one from each of two samples, are shared by both samples. Chao-Jaccard similarity score also corrects for the undersampling bias and thus is especially useful for comparing rich and incompletely sampled communities. The average pairwise Chao-Jaccard similarity score between all adenoid samples

was 0.26 (range 0.00-1.00, SD 0.26), indicating substantial compositional differences or beta-diversity between samples.

UniFrac analyses provided further evidence that there were large interpersonal variations in bacterial composition. Based on the weighted UniFrac distances that account for the abundance of taxa, the adenoid samples can be divided into at least five distinct groups, with each group having its unique bacterial composition. For example, group 1 was dominated by Proteobacteria, while group 2 and 3 were dominated by Firmicutes and Fusobacteria respectively (Figure 2).

For patient 84, we swapped the superior and inferior surfaces of adenoid independently and treated them as two separate samples through the entire data collection and analysis pipeline. In Figure 2, the superior and inferior samples of patient 84 are clustered together, indicating that the microbial communities that inhabited the superior and inferior surfaces of the adenoid were very similar. The fact that these two independently processed samples from one patient showed almost identical bacterial composition indicates that our data collection pipeline consisting of the genomic DNA extraction, 16S rRNA PCR amplification and 454 pyrosequencing was fairly robust.

Opportunistic pathogens

Because many of the abundant genera were known to contain potential pathogenic species, we further classified OTUs within these abundant genera. Using the software MEGAN and searching against the RDP database, we were able to identify many of the common upper respiratory infection pathogens including *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Staphylococcus aureus*, *Haemophilus influenzae, Moraxella catarrhalis* and *Neisseria*

meningitidis (Table 1). The most prevalent pathogen was *H. influenzae*. It was found in 70.6% of the samples and made up 10.5% of the total sequence reads. The next most prevalent pathogen was *S. aureus*, with a prevalence of 54.4% and an average relative abundance of 13.0%. This was followed by *S. pneumoniae, N. meningitidis* and *M. catarrhalis* with each having a prevalence of around 40%. However, they were much less abundant than *H. influenzae* and *S. aureus*. They only made up 0.2%, 0.4% and 2.3% of the total reads respectively. In some patients, the adenoid microflora was predominated by one single pathogenic species. For example, *S. aureus* constituted more than 90.0% of the reads in patient 59, 61, 90 and 95 while *S. pyogenes* made up more than 95.0% of the bacterial population in patient 96.

Core adenoid microbiota

We next investigated whether there was a core set of bacterial taxa shared between individuals. Only Streptococcus (genus) of Firmicutes (phylum) was ubiquitously present in all 67 individuals. When we relaxed the criterion and defined the core taxa as taxa that appeared in at least 80% of all individuals, we found three additional core phyla: Proteobacteria (95.6%), Fusobacteria (82.4%) and Actinobacteria (80.8%). We also found three additional core genera: Gemella (89.7%), Haemophilus (80.9%) and Fusobacterium (80.9%). At the 97% OTU level, only OTU 2242 (*Streptococcus mitis*, 83.8%) fit this criterion. The core taxa were also highly abundant, representing 89.4% and 42.4% of total sequences at the phylum and genus level respectively.

Co-occurrence of taxa

To investigate whether taxa were randomly distributed among individuals, we carried out co-occurrence analyses. We calculated the C-scores (observed) as described in (Stone and Roberts,

1990) and compared them to the C-scores of simulated bacterial communities (expected) using the software EcoSim. Significant difference between the observed C-score and the expected C-score would indicate that the taxon distribution pattern is not random.

When all taxa were compared, at the phylum level, the observed C-score was not significantly different from the random distribution (the null hypothesis). At the genus level, however, the observed C-score was significantly larger than the expected C-score ($p < 0.001$), suggesting possible segregation or competition among genera. Because of the large number of species present in our data, at the species level, we analyzed the distribution patterns only within a given genus. For the eleven most abundant genera, species within seven of them (Streptococcus, Fusobacterium, Haemophilus, Neisseria, Gemella, Moraxella and Prevotella) showed significant segregation or competition ($p < 0.05$).

We then carried out MINE analysis to search for individual OTU pairs that showed significant relationships. MINE is a powerful statistic tool for identifying and characterizing dependent relationships among hundreds of thousands of variable pairs. Our MINE analysis identified 3436 significant relationships (out of 117,370 pairs) between 485 OTUs that had at least 10 reads (FDR adjusted p value <0.05). When examining the 1,976 top scoring non-linear relationships (MIC-$\rho 2 > 0.2$, Supplementary Table 1), we observed a common segregation pattern in which a pair of OTUs tended not to coexist. For example, OTU 2242, classified as non-pathogenic *S. mitis* and one of the most abundant OTUs, showed "non-coexistence" relationship with several pathogens including *S. aureus* (Figure 3A), *H. influenza*, and *M. catarrhalis*. Similar "non-coexistence" relationship was also observed between other non-pathogenic Streptococcus species and pathogens, for example, between *Streptococcus intermedius* and *S. aureus* (Figure 3B), and

between *Streptococcus salivarius* and *S. pyogenes* (Figure 3C). Although much less frequently, we did observe linear relationship between OTUs, for example, between two *S. aureus* subspecies (Figure 3D), indicating that they might prefer to live in similar environment.

Adenoid microbiota is distinct from those of other human body sites

A previous study has shown that various human body sites harbor very different microbial communities (Costello et al., 2009). Therefore, we compared the overall bacterial composition of adenoid with those of other body sites and asked whether these adenoids harbored their own unique microbiota. We note, however, the adenoids we sampled here were removed by surgery and might contain bacterial populations different from those of healthy adenoids. We downloaded the 16S rRNA pyrosequences from the Costello *et al.* study that surveyed human microbiota across multiple body sites (skin, gut, hair, oral cavity, nostril and external auditory canal etc). Because slightly different regions of the 16S rRNA were sequenced in the Costello *et al.* study, we only used the overlapped region (~200bp) for comparison. Principal coordinate analysis and Unweighted Pair Group Method with Arithmetic mean (UPGMA) clustering based on UniFrac distances show distinct clustering by body habitats (Figure 4 and Supplementary Figure 3). Adenoid microbiota is overall mostly similar to oral microbiota, although some samples show similarity to the microfloras of nostril, skin and the external auditory canal.

On average, an adenoid was estimated to contain 355 bacterial species (Chao1 estimate based on 97% OTU). In terms of bacterial species richness, adenoid was on the same level as nostril and tongue but was less diverse than the gut (Supplementary Figure 2).

Discussion

With an average Good's estimator of 94.8%, we expect to have covered the vast majority of bacterial diversity on the 67 adenoids. The depth of sequencing coverage (4,105 reads/sample) should allow us to identify rare organisms that constitute ~ 0.05% of the community (0.05% x 4105 = 2 reads since we required at least 2 reads to identify a taxon). Mindful of potential inflated diversity estimates from short reads (Quince et al., 2009; Kunin et al., 2010), we employed a set of stringent criteria to remove low quality and artificial reads. Our 16S rRNA based survey revealed surprisingly large bacterial diversity on human adenoids. We identified a total of 3,121 species (97% OTUs) from 67 individuals.

With an extensive system of crypts and located at the junction of nose and throat, adenoids trap bacteria that pass through the upper airways. Therefore, many of these rare bacterial species we observed might simply be transient passengers. The high bacterial diversity is also consistent with the notion that adenoid plays a surveillance role for the immune system by providing early exposure of immune system to pathogens. Although we observed significantly reduced diversity in some adenoid samples, the overall alpha diversity of adenoid microbiota was surprisingly high and was comparable to those of nostril and oral microbiota of healthy individuals. This is remarkable considering that inflammatory diseases such as Crohn's disease and ulcerative colitis are typically associated with markedly reduced microbial diversity (Manichanh et al., 2006). It has also been shown that the diversity of nasopharyngeal microbiota was reduced in children with acute otitis media (Hilty et al., 2012).

The three most abundant phyla in the adenoid microbiota are Firmicutes, Proteobacteria and Fusobacteria. A previous study also found that these three phyla dominate in the adenoid of one patient (Liu et al., 2011), indicating this might represent a general pattern. Although the top five

most abundant phyla are identical to those of neighboring microbiota, their relative abundance are fairly different. Nasopharyngeal microbiota are dominated by Proteobacteria followed by Firmicutes and Bacteroidetes (Bogaert et al., 2011), while in oral cavity the three predominant phyla are Firmicutes, Proteobacteria and Bacteroidetes (Costello, Lauber et al. 2009). Brook et al cultured bacteria from adenoid tissues of 60 children and they found that the predominant aerobes were Streptococcus, Haemophilus, Staphylococcus and Moraxella, and the most abundant anaerobes were Peptostreptococcus, Prevotella and Fusobacterium (Brook et al., 2000). At the genus level, our survey showed a profile very similar to Brook *et al*.'s culture-based survey but distinct from those of nasopharynx (most abundant: Moraxella, Haemophilus and Streptococcus) and oral cavity (most abundant: Streptococcus, Veillonella and Prevotella). Accordingly, comparison at the species level (97% OTU) showed that adenoid microbiota is distinct from those of other body sites (Figure 4), supporting the idea that microbiota composition is mainly determined by the body habitats (Costello et al., 2009).

Like the microbiota of the other body sites, adenoid microbiota exhibited high level of interpersonal variability (beta-diversity). For example, at the phylum level, the percentage of Firmicutes varied from 1.2% to 99.9%. Based on the OTU composition, we observed at least five distinct types of adenoid microbiota, suggesting that there was no one single signature microflora that can be associated with adenoid inflammation. The underlying causes of the interpersonal variations (e.g. environmental factors, historical exposures or human genotypes) have been documented in other body sites (Alm et al., 1999; Zoetendal, 2001; Stewart et al., 2005; Ley et al., 2006; Khachatryan et al., 2008; Turnbaugh et al., 2009). Because we did not have the information on individual patients (e.g., age, gender, reason of surgery), we could not speculate on what was causing the interpersonal variations in adenoid microbiota and what was shaping the

different types of adenoid microbial communities. Recent surveys of nasopharyngeal microbiota in infants and children, however, indicated that season was associated with shifts in the bacterial community structure while other factors such as age, sex, day care use were not (Bogaert et al., 2011; Hilty et al., 2012).

Previous studies have suggested that there is an association between otitis media and chronic adenoidal infection. Our study supports the hypothesis that adenoid may serve as the potential bacterial source for middle ear infections. We found the common pathogens of otitis media (Haemophilus, Streptococcus, Moraxella, Staphylococcus) and/or their close relatives to be both prevalent and abundant in the adenoids we surveyed. This is consistent with the result of the previous culture-based survey of 60 diseased and healthy adenoids (Brook et al., 2000). Although potentially pathogenic bacteria are often isolated from the nasopharynx of healthy children, they are either transient or only constitute a minor part of the nasopharyngeal flora (Swidsinski et al., 2007). Our result is consistent with the idea that under certain conditions, the opportunistic pathogens in the normal nasopharyngeal microflora can grow to dominance and cause infectious diseases.

Our survey showed that although diseased adenoids were mostly dominated by one pathogen, adenoidal infections could be polymicrobial in nature. We detected the coexistence of multiple pathogens, all at significant levels, in several adenoid samples. One possible mechanism of the multiple overlapping infections was revealed by a spatial survey of adenoid microbiota using fluorescence in situ hybridization (FISH). It showed that different pathogens could coexist by occupying different locations of the adenoids (Swidsinski et al., 2007). For example, *H.*

influenzae was mainly found in the lymphoid tissue while Streptococcus was typically found in fissures.

One goal of human microbiome studies is to determine, for a given body site, whether there exists a core set of microbial species (Turnbaugh et al., 2007; Turnbaugh et al., 2009). It has been proposed that such core microbiota might be pivotal in maintaining the homeostasis and health (Sekelja et al., 2011). Recently, the existence of core gut microbiota has been confirmed (Rajilic-Stojanovic et al., 2009; Turnbaugh et al., 2010; Sekelja et al., 2011). Despite the large interpersonal variations, we were able to identify one core genus Streptococcus, which was present in all the adenoid samples and was also the most abundant genus overall.

It has been postulated that commensal bacteria can interfere with bacterial pathogens by competing for resources (e.g., nutrients and attachment sites) or producing bacteriocins that kill pathogens (Brook, 1999). Non-pathogenic alpha-hemolytic Streptococcus, Prevotella and Peptostreptococci have been show to be effective at interfering with pathogens that infect the upper respiratory tracts (Bernstein et al., 1993; Fujimori et al., 1996; Brook and Gober, 2000; Tano et al., 2000; Brook, 2003; Walls et al., 2003; Brook, 2005). Previous culture-based studies have demonstrated that viridans streptococci (in particular, *Streptococci mitis*, *salivarius* and *sanguis*) can prevent the colonization of a variety of pathogens including *H. influenza*, *S. pneumoniae* and *S. aureus* (Brook, 1999, 2005; Benninger et al., 2011). Results of our co-occurrence analysis are consistent with the bacterial interference hypothesis. Our MINE analysis showed that for example, *S. mitis*, one of the most abundant species in our survey, displayed statistically significant inverse relationships in the relative abundance with pathogens such as *H. influenza*, *S. aureus* and *M. catarrhalis*. Although habitat filtering (segregation due to differences

in the habitat, e.g., different human genotypes) and neutral processes (e.g., different historic exposures) can not be eliminated as the possible explanations of the observed segregation patterns, based on the results of previous studies, we think bacterial interference is the most plausible interpretation.

In conclusion, this study shows that human adenoid encompasses complex, diverse and highly variable bacterial communities. The open, non-targeted 16S rRNA survey has revealed patterns consistent with the bacterial interference and pathogen reservoir hypotheses, and will enhance our understanding of the relationship between adenoid microbiota and upper airway infections.

Experimental Procedures

Sample collection

Discarded adenoids were collected after surgery from 98 children undergoing adenoidectomy for diseases including hyperthrophic adenoids, chronic adenoiditis, chronic otitis media and obstructive sleep apnea. The adenoids were transported in sterile container on ice to the laboratory. The first 55 adenoids obtained were frozen at -20°C and thawed prior to sampling. The rest of the adenoids were sampled fresh. A sterile cotton swab was used to collect the sample from the surface of each adenoid. The swab was moved over the entire surface of the adenoid in order to obtain a representative sample. In order to explore local differences in bacterial flora, we obtained additional swab samples from the superior and inferior area separately from one adenoid (patient 84). The swabs were stored in a 1.5 ml centrifuge tubes at -20°C before DNA extraction.

DNA extraction

DNA was extracted from the swabs by bead-beating and phenol-chloroform extraction. Briefly, the cotton tip of each swab was cut off and placed into a 2 ml bead tube. 500 µL of buffer (10 mM Tris-HCl pH 7.5, 5 mM EDTA, 100 mM NaCl), 500 µL of phenol:chloroform:isoamyl alcohol (25:24:1), 200 µL of 10% SDS and 400 µL of 0.1 mm zirconia/silica beads were added to the tube. The tube was then homogenized on a Biospec mini beadbeater for 3 min. After the bead beating, the tube was centrifuged for 10 min at 15,000 g. The top aqueous layer was transferred to a new tube, extracted once with an equal volume of chloroform:isoamyl alcohol (24:1) and then ethanol precipitated. The DNA pellet was resuspended in 50 µL of 10 mM Tris-HCl pH 8.0. DNA samples were stored at −20 °C until needed.

Tag-PCR amplification of the V1-V2 regions of the bacterial 16S rRNA gene

The V1-V2 hypervariable regions of the 16S rRNA gene were PCR amplified from extracted DNA samples using two primers containing the universal sequences 27F (5'-AGRGTTTGATCMTGGCTCAG-3') and 534R (5'-TTACCGCGGCTGCTGGCAC-3') respectively. A unique 10-bp barcode was added to the 5' of the forward primer sequence to tag the samples. The PCR conditions used were 94°C for 3 min, followed by 30 cycles of 94 °C for 30 sec, 57°C for 45 sec and 72°C for 60 sec, with a final extension of 5 min at 72°C.

Pyrosequencing

16S rRNA amplicons from different samples were pooled in equal molar ratios, gel purified and then sequenced with Titanium chemistry on a 454 Life Science Genome Sequencer FLX platform at University of Virginia Department of Biology Genome Core Facility according to the

standard 454 protocol. The sequences have been deposited in the MG-RAST server (http://metagenomics.anl.gov/).

Sequence processing and OTU classification

Sequence reads were processed using the QIIME pipeline (Caporaso et al., 2010b). After sorted by barcodes into separate samples, reads were filtered using a set of stringent criteria to remove low quality reads. Reads were removed if they were shorter than 200 bp or longer than 550 bp, had an average Phred equivalent quality score less than 25, did not contain proper primer/barcode sequences, or had an ambiguous base call (N) in the sequences. Chimeric sequences were identified using two methods implemented in QIIME, Chimera Slayer (Haas et al., 2011) and BLAST (Altschul et al., 1990). Chimeric reads detected by both methods were removed. To further clean up the data from potential sequencing artifacts, we excluded OTUs that only contained a single sequence read (singletons).

OTUs at 97% identity cutoff were identified by Uclust (Edgar, 2010). The most abundant sequence of each OUT was chosen as the representative sequence, assigned a taxonomy using the Ribosomal Database Project (RDP) classifier (Wang et al., 2007) and aligned with each other using Pynast (Caporaso et al., 2010a). Alignment was then filtered to remove columns comprised of only gaps and known to be excessively variable using the mask file from Greengenes (DeSantis et al., 2006). The filtered alignment was then used to build a phylogenetic tree using FastTree (Price et al., 2010).

For genera that contained potential pathogens of upper respiratory infections, we further classified the OTUs within these genera using the phylotyping algorithm of MEGAN (Mitra et

al., 2011). For each genus, we downloaded high quality 16S rRNA sequences of cultured isolates of the genus from the RDP database and used them as the reference sequences. We then BLASTN searched the representative OTU sequences against the reference sequences and assigned OTU the species name of its top matches in the reference if their sequences were greater than 97% identical.

The human microbiota sequences from Costello *et al.* study (Costello et al., 2009) were downloaded from European Nucleotide Archive. They were processed and analyzed using the same protocol as described above.

Bacterial diversity estimation and community comparisons

We used Good's nonparametric coverage estimator (Good, 1953) to evaluate our sequencing effort. Coverage (C) of a sample size n was estimated by the formula, $C=1-N_1/n$, where N_1 denotes the number of classes observed exactly once (singletons).

Chao1 and Shannon index were calculated using QIIME. While Chao1 estimates the number of OTUs in each sample, Shannon diversity index measures both species richness (the number of different taxa) and evenness (the relative abundance of each taxon). For all the beta diversity related analyses, the sequence reads were rarefied to remove the sequencing effort heterogeneity. UniFrac distances based on the abundance of 97% OTUs were calculated and used for Unweighted Pair Group Method with Arithmetic mean (UPGMA) clustering and Principal Coordinate Analysis (PCoA). To measure the similarity between any two samples, the pairwise Chao-Jaccard estimator was calculated for all possible pairs using EstimateS (Colwell, 2009).

Co-occurrence analysis

We tested for nonrandom taxon co-occurrence patterns using Ecosim's co-occurrence module (Gotelli, 2000). The data set was organized in a matrix of presence and absence, in which each row represented a taxon and each column represented a patient. We calculated the C-score (the average of checkboard units for all pairs of taxa) and compared it to the C-scores simulated using null random matrices. If the observed C-score is significantly larger than the expected C-score, it indicates possible segregation of taxa. Conversely, if the observed C-score is significantly smaller than the expected C-score, it suggests possible aggregation or cooperative interactions.

To identify pairs of taxa that might be in involved in competition or cooperation, we carried out the Maximal Information-based Nonparametric Exploration (MINE) analysis (Reshef et al., 2011). The dependence of two OTUs was measured by the maximal information coefficient (MIC) calculated using the pair's relative abundance in all samples. To lower the computational cost, we limited the MINE analysis to OTUs with at least 10 sequence reads, effectively reducing the number of OTUs to 485. OTU pairs were first sorted by their MIC scores. Their false discovery rate (FDR) adjusted p-values were then computed using the method described in (Benjamini and Hochberg, 1995). A relationship was considered significant if the adjusted p-value was less than 0.05. Linear relationships were identified using ρ, the Pearson product-moment correlation coefficient. Non-linear relationships were uncovered using the non-linearity score MIC-$\rho 2$ as described in (Reshef et al., 2011).

Author contributions

MW, MS and BW conceived the project and designed the experiment. DUG, NTN, and EKA prepared the sample. TR performed data analyses. TR drafted the manuscript. MW revised the manuscript and supervised the work. All authors read and approved the final manuscript.

Acknowledgements

The authors declare not to have any financial or non-financial competing interests. This work was supported by National Institutes of Health Grant DC003166-S1.

References

Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C., Ceraj, I., Distel, D.L., and Polz, M.F. (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**: 551-554.

Alm, J.S., Swartz, J., Lilja, G., Scheynius, A., and Pershagen, G. (1999) Atopy in children of families with an anthroposophic lifestyle. *Lancet* **353**: 1485-1488.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005) Host-bacterial mutualism in the human intestine. *Science* **307**: 1915-1920.

Benjamini, Y., and Hochberg, Y. (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**: 289-300.

Benninger, M., Brook, I., Bernstein, J.M., Casey, J.R., Roos, K., Marple, B., and Farrar, J.R. (2011) Bacterial interference in upper respiratory tract infections: a systematic review. *American journal of rhinology & allergy* **25**: 82-88.

Bernstein, J.M. (1999) Waldeyer's ring and otitis media - The nasopharyngeal tonsil and otitis media. *Int J Pediatr Otorhinolaryngol* **49**: S127-S132.

Bernstein, J.M., Faden, H.F., Dryja, D.M., and Wactawski-Wende, J. (1993) Micro-ecology of the nasopharyngeal bacterial flora in otitis-prone and non-otitis-prone children. *Acta Otolaryngol (Stockh)* **113**: 88-92.

Bernstein, J.M., Sagahtaheri-Altaie, S., Dryja, D.M., and Wactawski-Wende, J. (1994) Bacterial interference in nasopharyngeal bacterial flora of otitis-prone and non-otitis-prone children. *Acta Otorhinolaryngol Belg* **48**: 1-9.

Bernstein, J.M., Faden, H.S., Scannapieco, F., Belmont, M., Dryja, D., and Wolf, J. (2002) Interference of nontypeable *Haemophilus influenzae* and *Moraxella catarrhalis* by *Streptococcus oralis* in adenoid organ culture: a possible strategy for the treatment of the otitis-prone child. *The Annals of otology, rhinology, and laryngology* **111**: 696-700.

Bernstein, J.M., Haase, E., Scannapieco, F., Dryja, D., Wolf, J., Briles, D. et al. (2006) Bacterial interference of penicillin-sensitive and -resistant *Streptococcus pneumoniae* by *Streptococcus oralis* in an adenoid organ culture: implications for the treatment of recurrent upper respiratory tract infections in children and adults. *The Annals of otology, rhinology, and laryngology* **115**: 350-356.

Bik, E.M., Long, C.D., Armitage, G.C., Loomer, P., Emerson, J., Mongodin, E.F. et al. (2010) Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J* **4**: 962-974.

Bluestone, C. (1999) *Eustachian tube function and dysfunction*. Hamilton, ON, Canada: BC Decker.

Bogaert, D., Keijser, B., Huse, S., Rossen, J., Veenhoven, R., van Gils, E. et al. (2011) Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS ONE* **6**: e17035.

Brook, I. (1999) Bacterial interference. *Crit Rev Microbiol* **25**: 155-172.

Brook, I. (2003) Effects of antimicrobial therapy on the microbial flora of the adenoids. *J Antimicrob Chemother* **51**: 1331-1337.

Brook, I. (2005) The role of bacterial interference in otitis, sinusitis and tonsillitis. *Otolaryngology-Head and Neck Surgery* **133**: 139-146.

Brook, I., and Gober, A.E. (2000) In vitro bacterial interference in the nasopharynx of otitis media-prone and non-otitis media-prone children. *Archives of Otolaryngology-Head & Neck Surgery* **126**: 1011-1013.

Brook, I., Shah, K., and Jackson, W. (2000) Microbiology of healthy and diseased adenoids. *Laryngoscope* **110**: 994-999.

Brown, M.V., and Fuhrman, J.A. (2005) Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquatic Microbial Ecology* **41**: 15-23.

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

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

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-336.

Colwell, R.K. (2009). EstimateS: Statistical estimation of species richness and shared species from samples. Version 8.2. [WWW document]. URL http://purl.oclc.org/estimates

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009) Bacterial community variation in human body habitats across space and time. *Science* **326**: 1694-1697.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K. et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069-5072.

Dhooge, I., Van Damme, D., Vaneechoutte, M., Claeys, G., Verschraegen, G., and Van Cauwenberge, P. (1999) Role of nasopharyngeal bacterial flora in the evaluation of recurrent middle ear infections in children. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **5**: 530-534.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M. et al. (2005) Diversity of the human intestinal microbial flora. *Science* **308**: 1635-1638.

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

Faden, H., Waz, M.J., Bernstein, J.M., Brodsky, L., Stanievich, J., and Ogra, P.L. (1991) Nasopharyngeal flora in the first three years of life in normal and otitis-prone children. *The Annals of otology, rhinology, and laryngology* **100**: 612-615.

Faden, H., Duffy, L., Wasielewski, R., Wolf, J., Krystofik, D., and Tung, Y. (1997) Relationship between nasopharyngeal colonization and the development of otitis media in children. Tonawanda/Williamsville Pediatrics. *The Journal of infectious diseases* **175**: 1440-1445.

Fekete-Szabo, G., Berenyi, I., Gabriella, K., Urban, E., and Nagy, E. (2010) Aerobic and anaerobic bacteriology of chronic adenoid disease in children. *Int J Pediatr Otorhinolaryngol* **74**: 1217-1220.

Fujimori, I., Hisamatsu, K., Kikushima, K., Goto, R., Murakami, Y., and Yamada, T. (1996) The nasopharyngeal bacterial flora in children with otitis media with effusion. *Eur Arch Otorhinolaryngol* **253**: 260-263.

Giongo, A., Gano, K.A., Crabb, D.B., Mukherjee, N., Novelo, L.L., Casella, G. et al. (2011) Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J* **5**: 82-91.

Good, I. (1953) The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika* **40**: 237-264.

Gotelli, N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology* **81**: 2606-2621.

Grice, E.A., Kong, H.H., Conlan, S., Deming, C.B., Davis, J., Young, A.C. et al. (2009) Topographical and temporal diversity of the human skin microbiome. *Science* **324**: 1190-1192.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G. et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21**: 494-504.

Hilty, M., Qi, W., Brugger, S.D., Frei, L., Agyeman, P., Frey, P.M. et al. (2012) Nasopharyngeal microbiota in infants with acute otitis media. *The Journal of infectious diseases* **205**: 1048-1055.

Karlidag, T., Demirdag, K., Kaygusuz, I., Ozden, M., Yalcin, S., and Ozturk, L. (2002) Resistant bacteria in the adenoid tissues of children with otitis media with effusion. *Int J Pediatr Otorhinolaryngol* **64**: 35-40.

Khachatryan, Z.A., Ktsoyan, Z.A., Manukyan, G.P., Kelly, D., Ghazaryan, K.A., and Aminov, R.I. (2008) Predominant Role of Host Genetics in Controlling the Composition of Gut Microbiota. *PLoS ONE* **3**.

Khoramrooz, S.S., Mirsalehian, A., Emaneini, M., Jabalameli, F., Aligholi, M., Saedi, B. et al. (2012) Frequency of *Alloicoccus otitidis*, *Streptococcus pneumoniae*, *Moraxella catarrhalis* and *Haemophilus influenzae* in children with otitis media with effusion (OME) in Iranian patients. *Auris Nasus Larynx* **39**: 369-373.

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

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

Ley, R.E., Backhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005) Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* **102**: 11070-11075.

Liu, C.M., Cosetti, M.K., Aziz, M., Buchhagen, J.L., Contente-Cuomo, T.L., Price, L.B. et al. (2011) The otologic microbiome: a study of the bacterial microbiota in a pediatric patient with chronic serous otitis media using 16SrRNA gene-based pyrosequencing. *Arch Otolaryngol Head Neck Surg* **137**: 664-668.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L. et al. (2006) Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**: 205-211.

Marchisio, P., Claut, L., Rognoni, A., Esposito, S., Passali, D., Bellussi, L. et al. (2003) Differences in nasopharyngeal bacterial flora in children with nonsevere recurrent acute otitis

media and chronic otitis media with effusion: implications for management. *Pediatr Infect Dis J* **22**: 262-268.

Mazmanian, S.K., Liu, C.H., Tzianabos, A.O., and Kasper, D.L. (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**: 107-118.

Mitra, S., Stark, M., and Huson, D.H. (2011) Analysis of 16S rRNA environmental sequences using MEGAN. *BMC Genomics* **12**.

Nistico, L., Kreft, R., Gieseke, A., Coticchia, J.M., Burrows, A., Khampang, P. et al. (2011) Adenoid reservoir for pathogenic biofilm bacteria. *J Clin Microbiol* **49**: 1411-1420.

Ordovas, J.M., and Mooser, V. (2006) Metagenomics: the role of the microbiome in cardiovascular diseases. *Curr Opin Lipidol* **17**: 157-161.

Perry, M., and Whyte, A. (1998) Immunology of the tonsils. *Immunol Today* **19**: 414-421.

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

Quince, C., Lanzen, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M. et al. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* **6**: 639-641.

Rajilic-Stojanovic, M., Heilig, H.G.H.J., Molenaar, D., Kajander, K., Surakka, A., Smidt, H., and de Vos, W.M. (2009) Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ Microbiol* **11**: 1736-1751.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S., McCulle, S.L. et al. (2011) Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* **108 Suppl 1**: 4680-4687.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J. et al. (2011) Detecting novel associations in large data sets. *Science* **334**: 1518-1524.

Sekelja, M., Berget, I., Naes, T., and Rudi, K. (2011) Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. *ISME J* **5**: 519-531.

Stewart, J.A., Chadwick, V.S., and Murray, A. (2005) Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *J Med Microbiol* **54**: 1239-1242.

Stone, L., and Roberts, A. (1990) The Checkerboard Score and Species Distributions. *Oecologia* **85**: 74-79.

Swidsinski, A., Goktas, O., Bessler, C., Loening-Baucke, V., Hale, L.P., Andree, H. et al. (2007) Spatial organisation of microbiota in quiescent adenoiditis and tonsillitis. *J Clin Pathol* **60**: 253-260.

Tano, K., Olofsson, C., Grahn-Hakansson, E., and Holm, S.E. (1999) In vitro inhibition of *S. pneumoniae*, nontypable *H. influenzae* and *M. catharralis* by alpha-hemolytic streptococci from healthy children. *Int J Pediatr Otorhinolaryngol* **47**: 49-56.

Tano, K., Grahn-Hakansson, E., Holm, S.E., and Hellstrom, S. (2000) Inhibition of OM pathogens by alpha-hemolytic streptococci from healthy children, children with SOM and children with rAOM. *Int J Pediatr Otorhinolaryngol* **56**: 185-190.

Tano, K., Hakansson, E.G., Holm, S.E., and Hellstrom, S. (2002) Bacterial interference between pathogens in otitis media and alpha-haemolytic Streptococci analysed in an in vitro model. *Acta Otolaryngol (Stockh)* **122**: 78-85.

Tomonaga, K., Kurono, Y., Chaen, T., and Mogi, G. (1989) Adenoids and otitis media with effusion: nasopharyngeal flora. *Am J Otolaryngol* **10**: 204-207.

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

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

Turnbaugh, P.J., Quince, C., Faith, J.J., McHardy, A.C., Yatsunenko, T., Niazi, F. et al. (2010) Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc Natl Acad Sci U S A* **107**: 7503-7508.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E. et al. (2009) A core gut microbiome in obese and lean twins. *Nature* **457**: 480-484.

Walls, T., Power, D., and Tagg, J. (2003) Bacteriocin-like inhibitory substance (BLIS) production by the normal flora of the nasopharynx: potential to protect against otitis media? *J Med Microbiol* **52**: 829-833.

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

Winther, B., and Innes, D.J. (1994) The human adenoid. A morphologic study. *Arch Otolaryngol Head Neck Surg* **120**: 144-149.

Winther, B., Gross, B.C., Hendley, J.O., and Early, S.V. (2009) Location of bacterial biofilm in the mucus overlying the adenoid by light microscopy. *Arch Otolaryngol Head Neck Surg* **135**: 1239-1245.

Wostmann, B.S., Larkin, C., Moriarty, A., and Bruckner-Kardoss, E. (1983) Dietary intake, energy metabolism, and excretory losses of adult male germfree Wistar rats. *Lab Anim Sci* **33**: 46-50.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M. et al. (2012) Human gut microbiome viewed across age and geography. *Nature* **486**: 222-227.

Zoetendal, E.G.A., A.D.L. Akkermans-van Vliet, W.M. Visser, J.A.G.M. de Vos, W.M. de (2001) The host genotype affects the bacterial community in the human gastrointestinal tract. *Microbial Ecology in Health and Disease* **13**: 129-134.

Tables and Figures

Table 1. The prevalence and relative abundance of common pathogens that cause upper airway infections

Pathogens	Prevalence (%)	Relative Abundance (%)	
		Average	Maximum
Haemophilus influenzae	70.6	10.5	74.7
Staphylococcus aureus	54.4	13.0	93.5
Moraxella catarrhalis	44.1	2.3	42.7
Neisseria meningitidis	41.2	0.4	8.1
Streptococcus pneumoniae	41.2	0.2	5.1
Streptococcus pyogenes	29.4	3.9	94.9

Figure 1. The bacterial species richness varied greatly between adenoid samples. The rarefaction analysis was performed by plotting the number of observed OTUs in each adenoid sample against the sequencing effort. Each line represents one of 68 adenoid samples. OTUs were defined using 97% similarity cutoff.

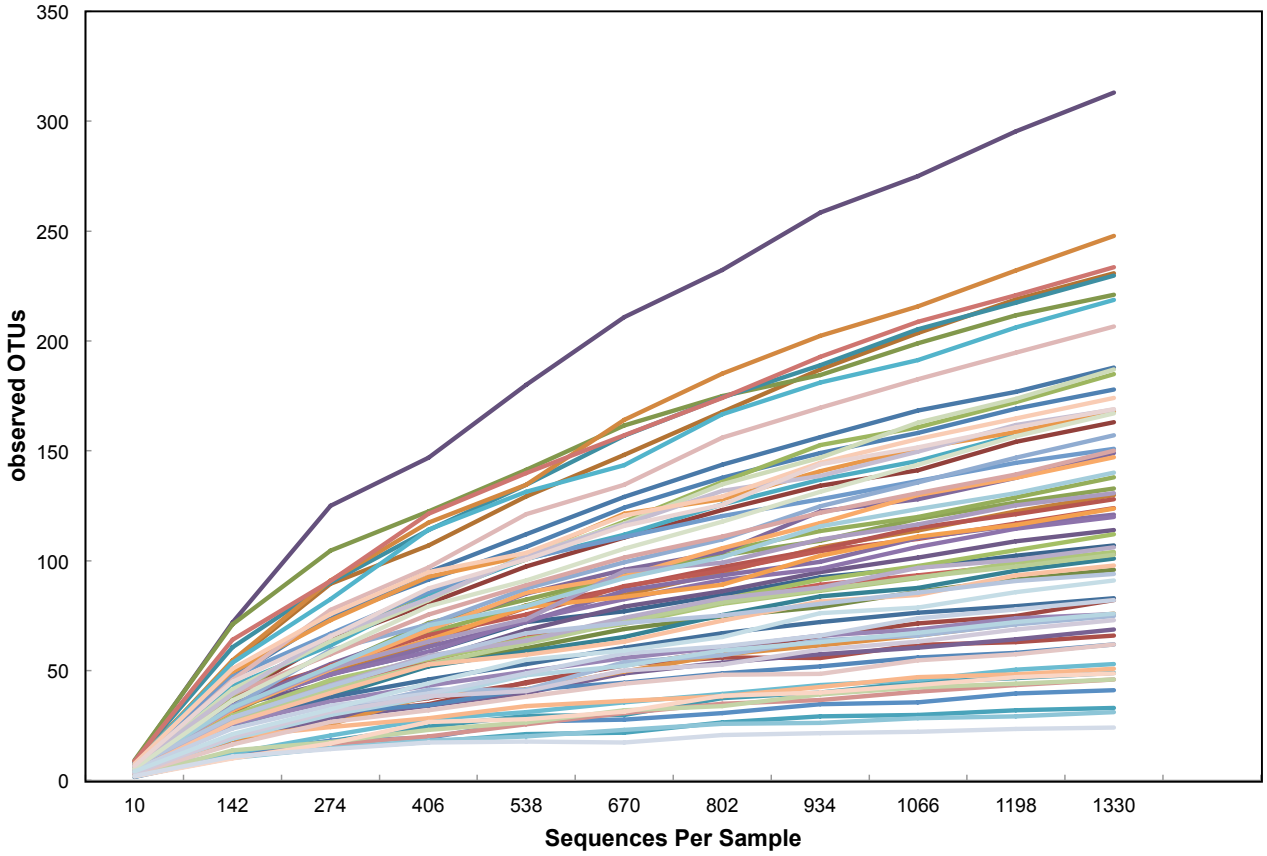


Figure 2. Adenoid microbiota composition varied greatly between adenoid samples. Taxonomic composition was broken down at the phylum level using Ribosomal Database Project classifier. Y-axis values represent the relative abundance of each phylum. Adenoid samples were clustered into five major groups with UPGMA of weighted UniFrac distances, with each group having a unique bacterial composition.

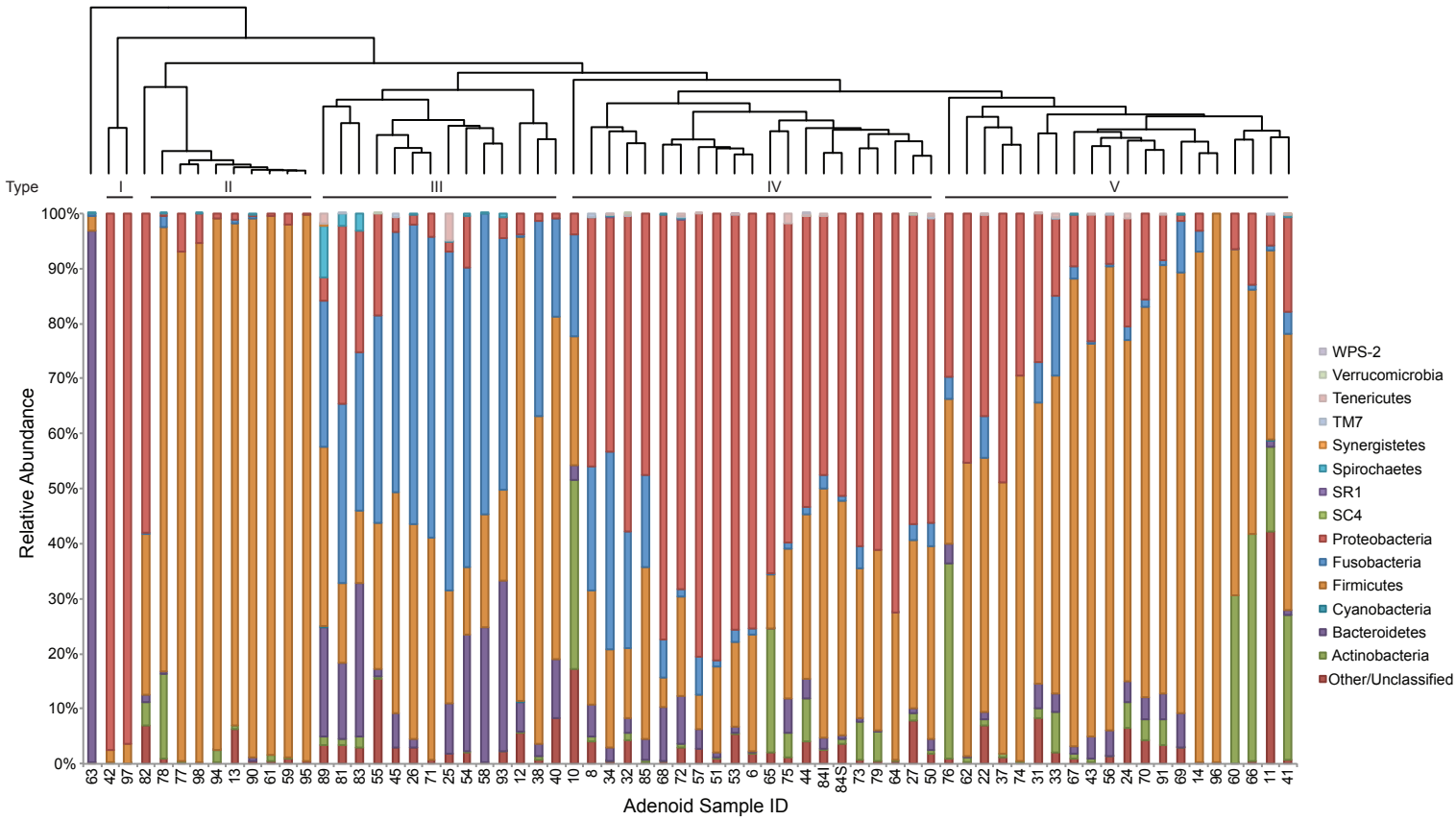


Figure 3. Coexistence and non-coexistence relationships between OTUs in the adenoid microbiota. Several non-pathogenic Streptococcus species display inverse relationships with common pathogens of upper airway infections in their relative abundance. A. between *S. mitis* (OTU 2242), one of the most abundant members of adenoid microbiota and *S. aureus* (OTU 4730), B. between *Streptococcus intermedius* (OTU 3426) and *S. aureus* (OTU 902), C. between *Streptococcus salivarius* (OTU 5198) and *Streptococcus pyogenes* (OTU 225). D. shows a linear relationship between two subspecies of *S. aureus* OTU 902 and OTU 5151. Each dot represents one adenoid sample. The *x* and *y* axes represent the relative abundance of the each OTU.

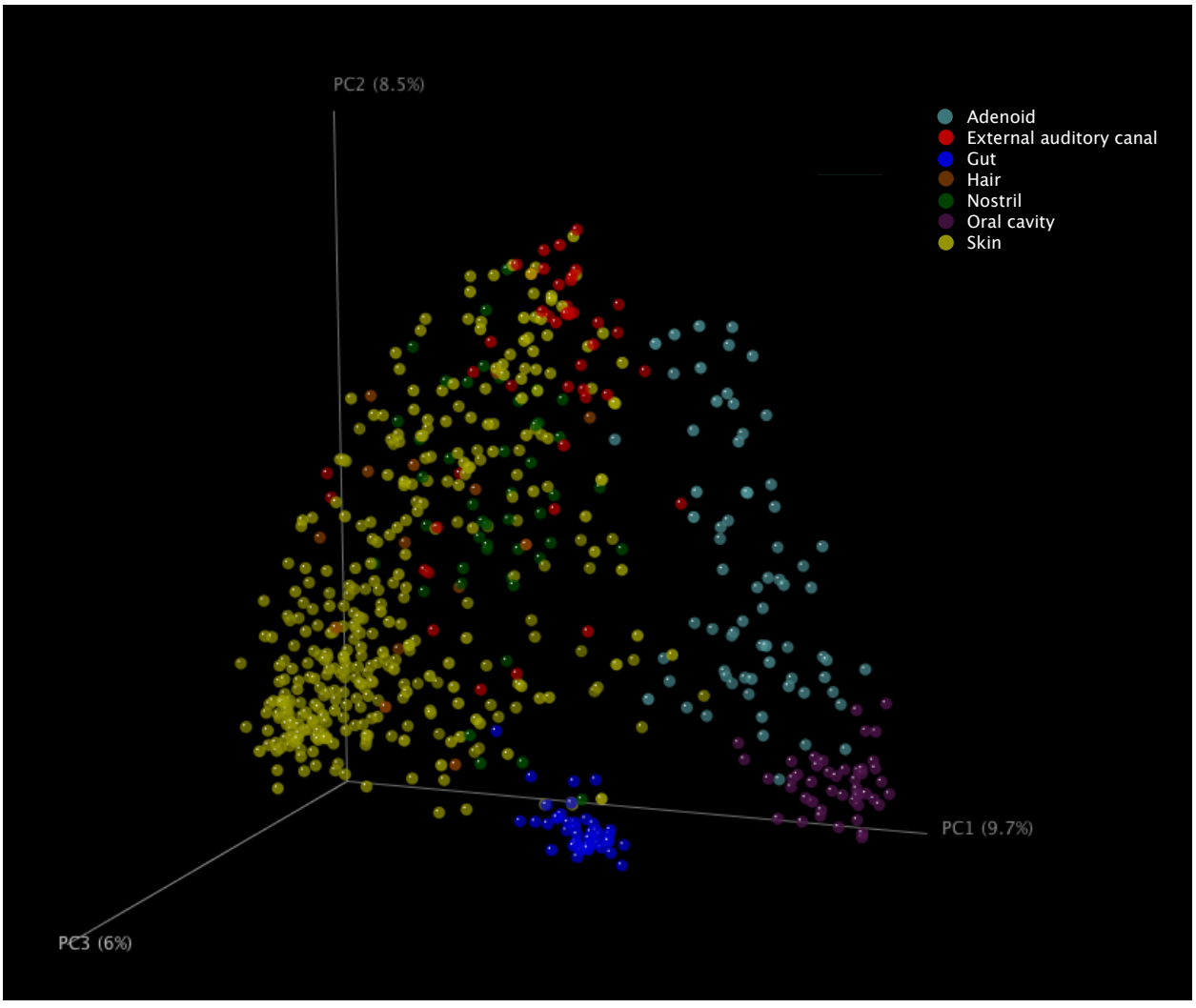
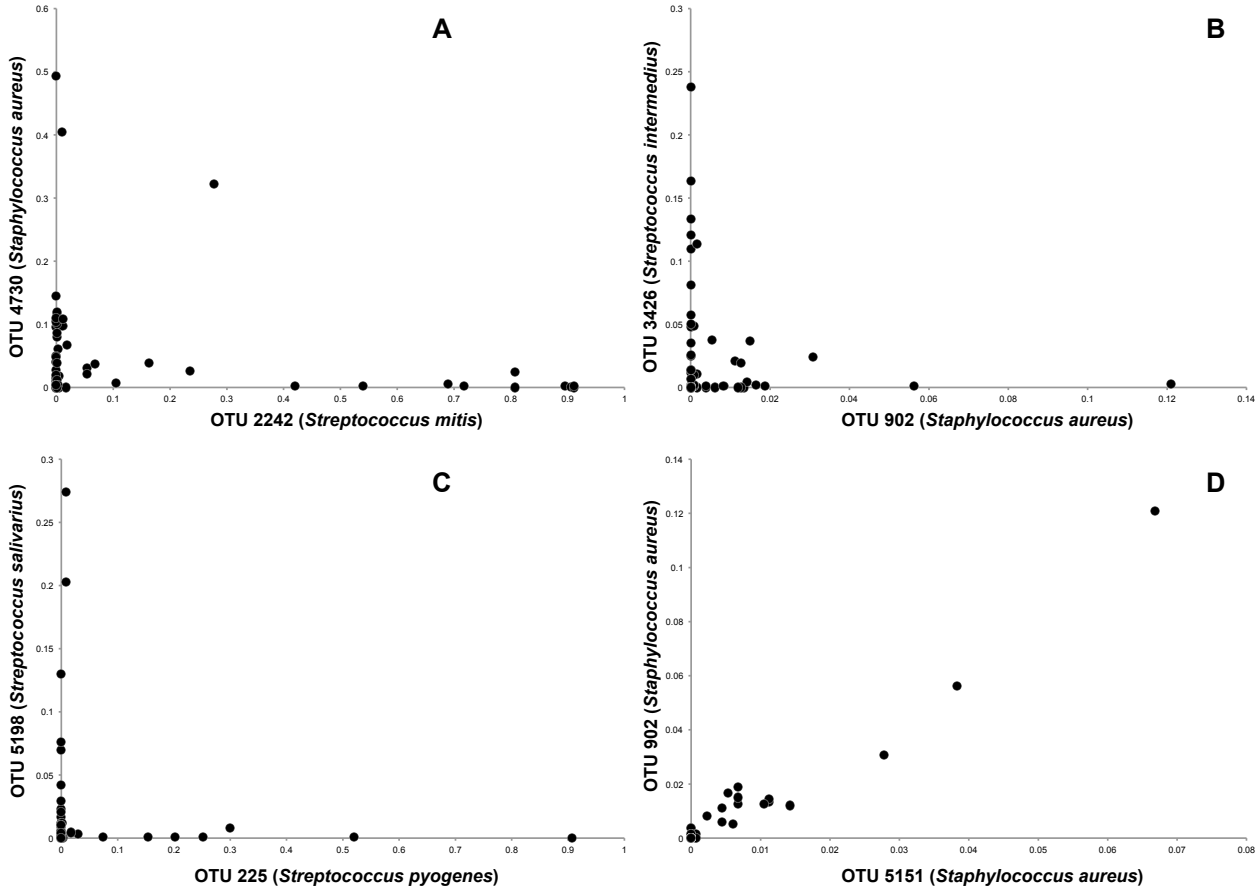
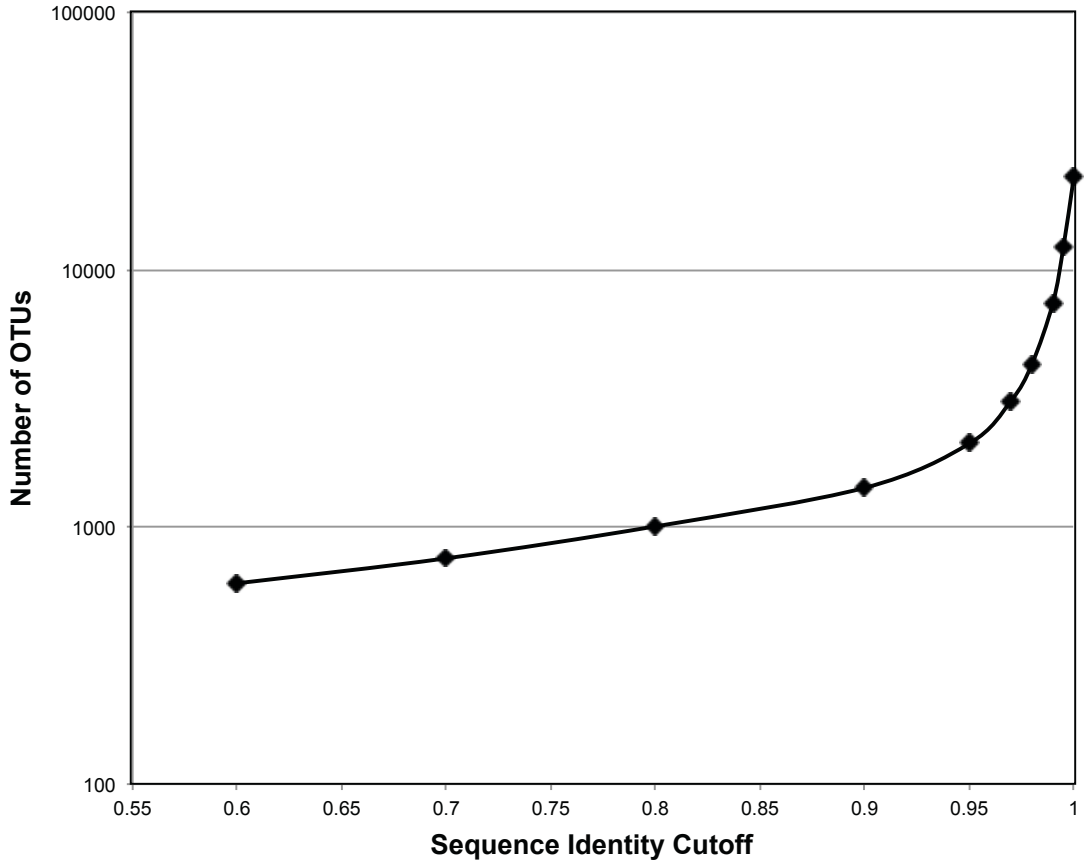


Figure 4. Adenoid microbiota is distinct from the microbiota of other human body sites. Principal coordinate analysis (PCoA) was performed based on the weighted UniFrac distance matrix. Each point represents a sample, colored by body site. The percentage of variation explained by each principal coordinate is indicated in parentheses.



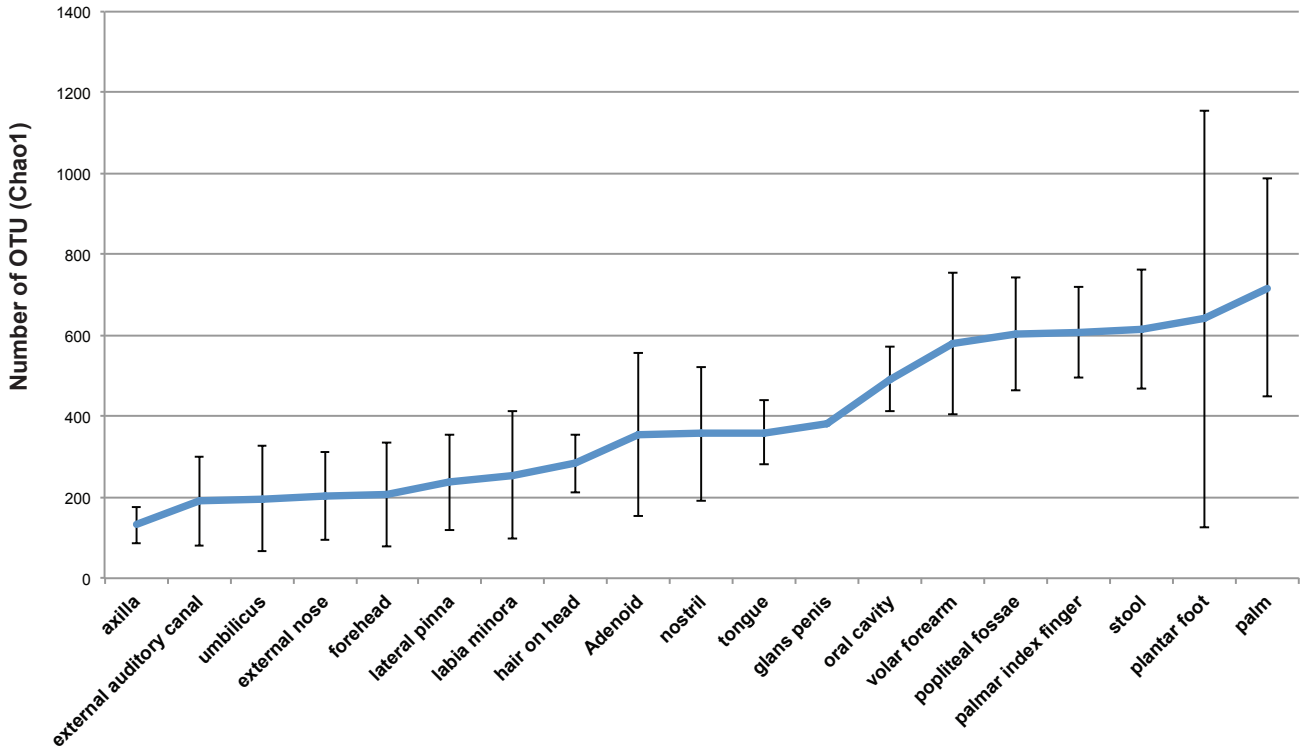
Supplementary information

Supplementary Figure 1. Number of OTUs as a function of sequence identity cutoff. A flare-up at 99% cutoff indicates that most of the bacterial diversity exists below the 97% identity cutoff that is used to approximate species.



Supplementary Figure 2. Bacterial species richness of adenoid and other human body sites.

The average number of OTUs for each body site was estimated by Chao1. Error bars represent +/- one standard deviation.



Supplementary Figure 3. Adenoid microbiota is distinct from the microbiota of other human body sites. UPGMA clustering was performed based on the weighted UniFrac distance matrix.

Supplementary Table 1. Pairs of OTUs that show significant non-linear relationships by the MINE analysis. The table was sorted by the non-linear score from the highest to the lowest.

Supplementary Figure 3 (5.8 MB) and Supplementary Table 1 (.xlsx) are available at:

http://onlinelibrary.wiley.com/wol1/doi/10.1111/1462-2920.12000/suppinfo

Appendix 2. PhyloCore: a phylogenetic approach to identifying core taxa in microbial communities

Formatted as a co-authored manuscript (Ren T and Wu M) in review at *Gene*

Abstract

Background

Identifying core microbiota is an important step for understanding the key components of microbial communities. Traditional approach that identifies core taxa at the OTU level ignores potential ecological coherence of higher rank taxa. There is a need to develop software that can systematically identify core taxa at and above the species level.

Results

Here we developed PhyloCore, an application that uses a phylogeny-based algorithm to identify core taxa at the proper taxonomic levels. It incorporates a number of features that users can set according to their needs. Using multiple gut microbiota as test cases, we demonstrate that PhyloCore is more powerful and flexible than OTU-based approaches.

Conclusions

PhyloCore is a flexible and fast application that identifies core taxa at proper taxonomic levels, making it useful to sequence-based microbial ecology studies. The software is freely available at http://wolbachia.biology.virginia.edu/WuLab/Software.html

Keywords: Core taxa, microbiota, OTU

Introduction

Core microbiota are defined as members shared by most microbial assemblages from similar habitats. It has been suggested that core taxa may play important roles in the function of the community (Shade and Handelsman, 2011). Thus, identifying core is a very useful step in microbial ecology studies. It provides valuable insights into what 'healthy' microbiota look like for a particular habitat and may also help in identifying keystone species in the community. The traditional approach used to identify core taxa is to find OTUs (operational taxonomic units) that are present in a large proportion of samples (Caporaso et al., 2011; Huse et al., 2012; Martínez et al., 2013). However, this method can only identify core taxa at the OTU level and ignores potential phylogenetic redundancy in microbial communities where multiple closely related OTUs are present.

Since closely related bacterial taxa can be ecologically interchangeable (Harvey and Pagel, 1991), it may be useful to consider phylogenetic relationships when identifying core taxa. Recent studies suggest that there is ecological coherence (members of a taxonomic group share common ecological traits that distinguish them from members of other taxonomic groups) among bacterial taxonomic ranks higher than the species (OTU) level (Lozupone and Knight, 2007; 2005; Philippot et al., 2010; 2009). These findings suggest that microbial cores might exist at taxonomic levels higher than species. Therefore, it is not surprising that traditional core identification method often fails to detect core OTUs. Higher taxonomic level cores have been identified in previous studies (Benson et al., 2010; Zhang et al., 2015). However, the cores were identified either manually or using in-house scripts that are not available to the public. There is a need to develop software that can systematically identify core taxa at and above the species level.

Here we present PhyloCore, an application that uses a phylogeny-based algorithm to identify core taxa at the proper taxonomic levels.

Material and methods

Algorithm

PhyloCore takes an OTU table that describes the presence/absence of each OTU in all samples and their taxonomic assignments, and optionally a phylogenetic tree of all OTUs. When a phylogenetic tree of all OTUs is provided, PhyloCore will use it to infer relationships between OTUs. In the absence of an OTU tree, PhyloCore will construct a tree using taxonomic information in the OTU table. OTU table and tree can be generated from 16S rRNA (or SSU rRNA) sequence data using a microbial ecology analysis software such as QIIME (Caporaso et al., 2010). To identify the core taxa, PhyloCore starts at the root node and traverses the whole tree in breadth-first order. For each internal node, PhyloCore calculates a prevalence value, defined as the cumulative presence of all its descendant OTUs. For example, in Figure 1 the prevalence of the internal node B is 2/3, because its two descendant OTUs (OTU 1 and OTU 2) appear in 2 out of 3 samples. For each leaf node or OTU, (e.g., OTU 4 in Fig. 1), PhyloCore calculates its prevalence value in all samples.

PhyloCore initially stores any node *i* that has a prevalence value greater than a user supplied threshold as a core taxon. However, if PhyloCore finds a descendant of node *i* (e.g., node *j*) that also passes the threshold, then node *i* is replaced by node *j* in the stored core node list. For example, using a 0.5 threshold, node A is initially defined as a core node. However, when PhyloCore moves down the lineage and finds that node B also qualifies, node A is replaced by node B in the core node list. This will guarantee that for a particular lineage only the core at the

lowest possible taxonomic level is identified, assuming that cores at lower taxonomic levels are more informative than those at higher levels in revealing the functions of the core that are important for the community.

Once a list of core nodes is identified, PhyloCore assigns taxonomy to each core node. If the core is an OTU, its taxonomy is used directly. For an internal core node, it is done by finding the lowest common taxonomy of all its descendant OTUs. Take node B for example, the lowest common taxonomy of OTU 1 and OTU 2 is family Lachnospiraceae. Therefore, Lachnospiraceae is assigned to node B.

Dataset can be imbalanced when sampling among groups is uneven. For example, in Figure 1 group I has two samples while group II has only one sample. This will result in a bias towards group I if all samples are treated equally. In this case, weighted core taxa can be identified. The prevalence of each node P_{weighted} will be calculated based on the formula below, giving each group (but not each sample) the same weight:

$$\boldsymbol{P_{weighted}} = \frac{1}{N}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}(P_i)$$

where P_i is the prevalence in i'th group, and N is the total number of groups.

Features

PhyloCore is coded in Perl and allows user to specify:

a) A prevalence threshold. A node is considered a core node if its prevalence is above the threshold.

b) An abundance threshold. OTUs with abundances lower than the threshold in a sample will be considered absent.

c) A sample ID list (with or without group information). Only samples in the list will be used in core identification. The group information, if provided, will be used to identify the weighted core nodes. This feature enables users to identify cores in a subset of the population (e.g., a specific group).

Results and Discussion

To demonstrate the use of PhyloCore, we first identified the core gut microbiota in mammals. The dataset contains 16S rRNA gene sequenced from 85 individuals belonging to 6 mammalian orders (Ley et al., 2008). Using QIIME, a 16S rRNA tree and an OTU table were generated and used as input files to PhyloCore. Prevalence threshold was set at 0.8 and samples were weighted. PhyloCore identified many core gut microbiota among different mammalian lineages (Figure 2). In comparison, traditional OTU-based core identification (as implemented in QIIME package) found only one OTU-level core (OTU 2209: family RFP12) in Perissodactyla, which was identified by PhyloCore as well. We found one order-level (Bacteroidales) and two family-level (Ruminococcaceae and Lachnospiraceae) core taxa that are shared among the mammals studied, suggesting that these taxa were likely present in the last common ancestor of mammals and are important in the codiversification of the gut microbiota and the mammalian hosts. As expected, the core hierarchy generally parallels the hierarchy of the host species phylogeny. In other words, if a microbial taxon *a* is a core of a host taxon *b*, then descendants of *a* are also cores of the descendants of *b*. For example, Order Bacteroidales is a core for Euarchontoglires. Genera YRC22 and Prevotella within Bacteroidales are cores of Rodentia and Primates, respectively (Figure 2). The parallelism in hierarchy should be perfect when the core is defined as being

present in all samples (i.e., using a prevalence threshold of 1.0). Because a prevalence threshold of 0.8 was used here, the core hierarchy breaks down in some lineages. For example, Family Ruminococcaceae was identified as a core for Placentalia, but neither Ruminococcaceae nor its descendants were cores for Carnivora (Figure 2).

We also tested PhyloCore on 16S rRNA sequences from two human microbiome studies (Caporaso et al., 2011; Yatsunenko et al., 2012). The Yatsunenko et al. dataset contained 528 samples collected from healthy children and adults from Amazonas of Venezuela, rural Malawi and US metropolitan areas. The Caporaso et al. dataset encompassed 467 samples collected from two individual over 1 year period. 16S rRNA trees and OTU tables generated by QIIME were used as input files to PhyloCore. Yatsunenko et al. dataset contained 45,595 OTUs, and it took PhyloCore (Python version) 4 minutes to run on a Macbook (1.6 GHz Intel Core i5 processor and 8 GB of memory). At the 0.9 prevalence cutoff, QIIME found two OTU-level core taxa (OTU 1: Dorea and OTU 5: Blautia). Besides these two OTUs, PhyloCore identified additional core gut microbiota at higher taxonomic levels: one order (Bacteroidales), two families (Coriobacteriaceae and Veillonellaceae) and two genera (Faecalibacterium and Streptococcus). Caporaso et al. dataset contained 4,926 OTUs, and it took PhyloCore 25 seconds to run on the same computer. At the 0.9 prevalence cutoff, QIIME found 15 OTU-level core taxa (12 Bacteroides, 1 Roseburia, 1 Phascolarctobacterium and 1 unclassified Ruminococcaceae). In comparison, PhyloCore identified two more genus-level (Faecalibacterium and Blautia) core taxa in addition to the 15 core OTUs.

The confidence of core identification depends on the sample size. For example, using the same prevalence threshold, we would place more confidence in core taxa that are identified from 1,000

samples than those identified from 10 samples. It is therefore important that sufficient number of samples are included in the study. As a general guideline, the smaller the number of samples used in a study, the more stringent prevalence threshold should be applied. Ultimately, the users should decide what is a proper prevalence threshold for a given sample size. At the minimum, the users should report the sample size and the prevalence threshold used in the core analysis.

Based on the premise that ecological coherence can exist at higher taxonomic levels, we think it is useful to identify core microbiota at and above the species level. However, it is worth pointing out that by no means it implies that members of core taxa identified by PhyloCore all have the same functions, as many studies have demonstrated that closely related bacterial species do not have completely overlapping functions (Cordero et al., 2012; Youngblut et al., 2013). Instead, it suggests that functions shared by core taxa might be important for the function of the community. Although the correlation between 16S rRNA tree and ecological functions is not perfect, many studies have demonstrated an overall strong correlation that should be useful in predicting species function from phylogeny (Langille et al., 2013; Snel et al., 1999; Zaneveld et al., 2010).

Conclusion

We have developed PhyloCore, an application that uses phylogeny to identify core taxa in microbial communities. It has been suggested that core microbiota exist at the functional rather than the taxonomic level (Consortium, 2013; Turnbaugh et al., 2009). However, these two alternative hypotheses are not mutually exclusive. It is conceivable that core functions are shared by the phylogenetic core taxa and therefore they represent the two aspects of the same microbiota core. Having PhyloCore will help us further test this theory.

Availability and requirements

Project name: PhyloCore

Project home page: http://wolbachia.biology.virginia.edu/WuLab/Software.html

Operating system(s): Unix/Linux (Perl and Python versions); Mac OS (Python version)

Programming language: Perl or Python 2

Other requirements: Bioperl 1.5.2 or later, or Biopython and Numpy

License: GNU GPL

Any restrictions to use by non-academics: None

Conflict of interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

MW and TR conceived the project and designed the software. TR implemented and tested the software. TR drafted the manuscript. MW revised the manuscript and supervised the work. All authors read and approved the final manuscript.

Acknowledgements

This work has been supported by the National Institutes of Health [5R01GM108501 to MW].

References

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., Kachman, S.D., Moriyama, E.N., Walter, J., Peterson, D.A., Pomp, D., 2010. 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. 107, 18933–18938.

Caporaso, J.G., Kuczynski, J., Stombaugh, J.I., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Meth 7, 335–336.

Caporaso, J.G., Lauber, C.L., Costello, E.K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J.I., Knights, D., Gajer, P., Ravel, J., Fierer, N., Gordon, J.I., Knight, R., 2011. Moving pictures of the human microbiome. Genome Biology 12, R50.

Consortium, T.H.M.P., 2013. Structure, function and diversity of the healthy human microbiome. Nature 486, 207–214.

Cordero, O.X., Wildschutte, H., Kirkup, B., Proehl, S., Ngo, L., Hussain, F., Le Roux, F., Mincer, T., Polz, M.F., 2012. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. Science 337, 1228–1231.

Huse, S.M., Ye, Y., Zhou, Y., Fodor, A.A., 2012. A core human microbiome as viewed through 16S rRNA sequence clusters. PLoS ONE 7, e34242.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Thurber, R.L.V., Knight, R., Beiko, R.G., Huttenhower, C.,

2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31, 814–821.

Ley, R.E., Hamady, M., Lozupone, C.A., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., Gordon, J.I., 2008. Evolution of Mammals and Their Gut Microbes. Science 320, 1647–1651.

Lozupone, C.A., Knight, R., 2007. Global patterns in bacterial diversity. Proc. Natl. Acad. Sci. U.S.A. 104, 11436–11440.

Lozupone, C.A., Knight, R., 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl. Environ. Microbiol. 71, 8228–8235.

Martínez, I., Muller, C.E., Walter, J., 2013. Long-Term Temporal Analysis of the Human Fecal Microbiota Revealed a Stable Core of Dominant Bacterial Species. PLoS ONE 8, e69621–12.

Philippot, L., Andersson, S.G.E., Battin, T.J., Prosser, J.I., Schimel, J.P., Whitman, W.B., Hallin, S., 2010. The ecological coherence of high bacterial taxonomic ranks. Nature Reviews Microbiology 8, 523–529.

Philippot, L., Bru, D., Saby, N.P.A., Cuhel, J., Arrouays, D., Simek, M., Hallin, S., 2009. Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of the 16S rRNA bacterial tree. Environ Microbiol 11, 3096–3104.

Reis, dos, M., Inoue, J., Hasegawa, M., Asher, R.J., Donoghue, P.C.J., Yang, Z., 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proc. Biol. Sci. 279, 3491–3500.

Shade, A., Handelsman, J., 2011. Beyond the Venn diagram: the hunt for a core microbiome. Environ Microbiol 14, 4–12.

Snel, B., Bork, P., Huynen, M.A., 1999. Genome phylogeny based on gene content. Nature

Genetics 21, 108–110.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., Egholm, M., Henrissat, B., Heath, A.C., Knight, R., Gordon, J.I., 2009. A core gut microbiome in obese and lean twins. Nature 457, 480–484.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C.L., Clemente, J.C., Knights, D., Knight, R., Gordon, J.I., 2012. Human gut microbiome viewed across age and geography. Nature 486, 222–227.

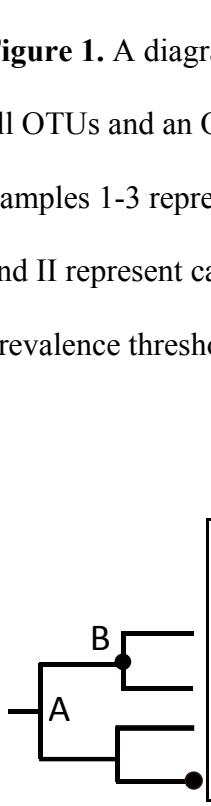
Youngblut, N.D., Shade, A., Read, J.S., McMahon, K.D., Whitaker, R.J., 2013. Lineage-specific responses of microbial communities to environmental change. Appl. Environ. Microbiol. 79, 39–47.

Zaneveld, J.R., Lozupone, C.A., Gordon, J.I., Knight, R., 2010. Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. Nucleic Acids Research 38, 3869–3879.

Zhang, J., Guo, Z., Xue, Z., Sun, Z., Zhang, M., Wang, L., Wang, G., Wang, F., Xu, J., Cao, H., Xu, H., Lv, Q., Zhong, Z., Chen, Y., Qimuge, S., Menghe, B., Zheng, Y., Zhao, L., Chen, W., Zhang, H., 2015. A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. The ISME Journal 1–12.

Figures

Figure 1. A diagram illustrating the criteria of core node identification. A phylogenetic tree of all OTUs and an OTU table describing the presence/absence each OTU are required as input. Samples 1-3 represent microbial communities (e.g. gut microbiota of different hosts). Group I and II represent categories of microbial community (e.g. from male or female hosts). Using a prevalence threshold of 0.5, internal node B and OTU 4 are identified as core nodes.



	Group I		Group II	Taxonomy
	Sample1	Sample2	Sample3	
OTU 1	1	0	0	Clostridiales; Lachnospiraceae; Blautia
OTU 2	0	1	0	Clostridiales; Lachnospiraceae; Dorea
OTU 3	1	0	0	Clostridiales; Ruminococcaceae; Ruminococcus
OTU 4	1	0	1	Clostridiales; Ruminococcaceae; Oscillospira

Prevalence threshold: 0.5 ●: Core node 1: present 0: absent

Figure 2. Core microbial taxa identified in different mammalian lineages using a prevalence cutoff of 0.8. P: phylum, O: order, F: family, G: genus. The mammalian tree topology was derived from (Reis et al., 2012). A 16S rRNA tree and an OTU table made with QIIME were used as input files to PhyloCore.