DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Disentangling the Coevolutionary Histories of Animal Gut Microbiomes

Citation

Sanders, Jon G. 2015. Disentangling the Coevolutionary Histories of Animal Gut Microbiomes. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:17463127

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Disentangling the coevolutionary histories of animal gut microbiomes

A dissertation presented

by

Jon Gregory Sanders

to

The Department of Organismic and Evolutionary Biology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Organismic and Evolutionary Biology

Harvard University
Cambridge, Massachusetts

April, 2015

cc 2015 – Jon G. Sanders

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Professor Naomi E. Pierce
Professor Peter R. Girguis

Jon G. Sanders

Disentangling the coevolutionary histories of animal gut microbiomes

ABSTRACT

Animals associate with microbes in complex interactions with profound fitness consequences. These interactions play an enormous role in the evolution of both partners, and recent advances in sequencing technology have allowed for unprecedented insight into the diversity and distribution of these associations. However, our understanding of the processes generating those patterns remains in its infancy. Here, I explore variation in microbiomes across two animal lineages—ants and mammals—to tease apart the role of these process in the evolution of gut microbiota. First, I explore patterns of phylogenetic correlation in gut microbiota of herbivorous *Cephalotes* ants and hominid apes. By examining the sensitivity of phylogenetic correlation to analytical parameters, I show that these outwardly similar patterns are likely to be the result of very different processes in each host lineage. Next, I examine in more depth the interacting effects of diet and phylogeny on the structure of baleen whale microbiomes. Whales consume a diet that differs dramatically from that of their closest extant relatives, the herbivorous artiodactyls. I use a combination of marker gene and shotgun metagenomic sequencing to show that a phylogentically conserved host trait, the multichambered gut, leads to functional and taxonomic similarities of whale gut microbiomes to those of their herbivorous ancestors via the fermentation of animal polysaccharides in the exoskeletons of their prey. Finally, I return to ants to examine how major shifts in the nature of gut microbial association correspond to host ecology. Using measures of absolute bacterial abundance, rather than diversity, I test the hypothesis that evolution of symbiosis with microbes has facilitated ants' dominance of tropical rainforest canopies. Surprisingly, I find differences in the abundance of gut bacteria in different ant lineages that span many orders of magnitude, suggesting that evolutionary transitions in the functional role of symbiosis in this animal lineage correspond not only to changes in the diversity of these associations, but to changes in kind. The results of these studies help to clarify the roles of history and selection in structuring animal gut microbiota, hinting that the interaction of these factors may fundamentally differ between animal lineages.

Disentangling the coevolutionary histories of animal gut microbiomes

TABLE OF CONTENTS

Acknowledgments **vi**

Introduction **1**

Chapter 1 **10**
Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes

Chapter 2 **41**
The whale gut microbiome digests animal prey using pathways common in terrestrial herbivores

Chapter 3 **63**
Gut bacterial densities help to explain the relationship between diet and habitat in rainforest ants

Appendix 1 **93**
Chapter 1 Supplemental

Appendix 2 **116**
Chapter 2 Supplemental

Appendix 3 **126**
Chapter 3 Supplemental

Appendix 4 **136**
Cladescan: a program for automated phylogenetic sensitivity analysis

References **139**

ACKNOWLEDGMENTS

Evolution is what happens.

There is something profound in this, the accidental grace of happenstance, that I find unfailingly compelling. I had always intended this thesis to be a tribute to the serendipity of cooperation. That it be so as much in execution as in topic is fitting.

It also means that I owe an enormous debt to a tremendous number of people.

First and foremost to my coadvisors, Naomi Pierce and Peter Girguis, for having the unending faith in (and patience with) me to find my own way through this journey. You have given me the chance to explore quite literally the deepest depths and the darkest forests in the farthest corners of this incredible planet in search of my questions. You have been friends and mentors, and will always remain an inspiration. I am incredibly grateful.

To my committee members, Andrew Knoll and Eric Alm; and to previous committee members Chris Marx, Anne Pringle, and Peter Turnbaugh; for guidance, support, and direction.

To a long list of collaborators and mentors at Stanford, Harvard, and elsewhere. I would not have made it here without rediscovering a love for biology at Hopkins Marine Station; thanks to Stephen Palumbi and George Somero, and the many other faculty, staff, and students there who started me on this road. I am incredibly fortunate to have shared an office with Daniel Kronauer, who shared with me the secrets of the *Cephalotes* phylogeny, as well as the secrets to taking their photographs. I must also thank Daniel for introducing me to Scott Powell, who has been an extraordinary friend, mentor, and guide to the wonderful world of turtle ants. For the gift of ants I also owe a profound debt to Jacob Russell and Corrie Moreau, whose work was an early inspiration, and who I am now lucky to call (along with their lab members Yi Hu, Piotr Lukasik, Ben Rubin) valued friends and collaborators. I am very, very grateful to have crossed paths with Megan

Frederickson, who has been an invaluable source of insight and support over many years. Thanks especially to Annabel Beichman, to whom I owe an entire chapter of my thesis. Thanks also to John Wertz, Joe Roman, Colleen Cavanaugh, Margaret McFall-Ngai, Ned Ruby, Rob Knight, Lawrence David, and Heraldo Vasconcelos.

To the fantastic members of the Pierce and Girguis labs, past and present. I have benefitted enormously from the friendship, intellectual stimulation, and moral support of both of these groups of people; I don't know how most people get by with just one lab. I must especially thank Heather Olins, Kiana Frank, Jenny Delaney, Stephanie Hillsgrove, and Chris DiPerna of the Girguis lab; and Gabe Miller, Leonora Bittleston, James Crall, Sarah Kocher, Maggie Starvish, and Petra Kubikova of the Pierce lab. In particular, I am deeply grateful for the friendship and support of Roxanne Beinart and Chris Baker, each of whom has been a constant and invaluable presence in my life since I arrived here.

To the many other people in OEB, the MCZ, and Harvard who have been sources of inspiration, assistance, and encouragement; especially Shane Campbell-Staton, Shelbi Russell, Stefan Cover, Jignasha Rana, and Chris Preheim.

To Millstone Coop and the extraordinary people I met there; Alex, Chandra, Sara, Mary, Benny (and of course the chickens). To Wes, David, Megan, Anna, Gil, Darcy, Katy, Cara, Laure, and Danica, friends through so many miles and so many years.

To Megan Jensen, my best friend of all.

To Lina Arcila Hernandez, my incredible partner, who is the most tenacious, genuine, and kind person I have ever met.

And of course, of course, to my family: Mom, Dad, Katie, Grandma Whiteman, and Grandma and Grandpa Watson. They have supported and encouraged me beyond what anyone could ever hope for or imagine. One cannot choose one's family, but it is the greatest serendipity of all that I ended up with this one.

INTRODUCTION

It is now widely appreciated, in both the academy and the public consciousness, that we are not simply the unitary organisms encoded by our nuclear genomes (McFall-Ngai *et al.* 2013). For humans, it is estimated that less than one percent of the total gene diversity found in or on our physical corpus is encoded by one of our 23 nuclear chromosomes (Qin *et al.* 2010). Just one in ten of the cells in that corpus harbor our nuclear genome. The vast majority of the remainder are bacteria, the bulk of them residing in the gut. We know that these symbiotic bacteria can have enormous effects on the phenotype of the host (McFall-Ngai *et al.* 2013)—and, consequently, on the realized fitness of those 23 human chromosomes.

But while the patterns of inheritance of nuclear DNA through the animal germ line are generally well characterized, the rules governing the trans-generational association between animal cells and the overwhelming gene diversity encoded within bacterial cells remain poorly understood. Despite profound advances in the technology available to characterize diverse microbial communities, the patterns of diversity even among some of the best-characterized host-associated microbiota remain frustratingly subject to differences in interpretation (Ochman *et al.* 2010; Moeller *et al.* 2012; Degnan *et al.* 2012; Sanders *et al.* 2013). To truly understand the dance of these more and less intimately-associated genomes across millions of years, their effect on and how they are affected by evolution of the host organism—that is to say, us—we must first learn how to interpret the patterns of their extant distributions with greater fidelity.

The phenotypic effects of animal microbiomes

Much of the growing interest in animal microbiota can be attributed to an increasing appreciation for the breadth of their effects on host phenotype. The best known of these effects are as sources of metabolic novelty: the diversity of pathways encoded in microbial genomes dwarfs those encoded endogenously in eukaryotes. Close symbioses with microbes can thus permit animals access to environments and energy sources wildly disparate from those of their non-symbiotic (or, rather, *differently*-symbiotic) ancestors. The most extreme of these metabolic partnerships may be the chemoautotrophic associations of hydrothermal vents (Cavanaugh *et al.* 1981; Dubilier *et al.* 2008), where many lineages of invertebrate animals have evolved associations with bacteria that permit them to use geochemical energy to fix carbon dioxide at enormous rates, forming the base of a food chain largely independent of solar energy. Microbial symbionts also play key roles in facilitating heterotrophic access to difficult reduced carbon sources such as lignocellulose, with plant-feeding animals ranging from ruminant and non-ruminant mammals (Pope *et al.* 2010; Zhu *et al.* 2011) to marine wood-boring bivalves (Distel & Roberts 1997) to termites (Breznak & Brune 1994), beetles (Geib *et al.* 2008), and ants (Pinto-Tomas *et al.* 2009; Suen *et al.* 2010; Aylward *et al.* 2012) all depending to some degree on microbial communities to digest intransigent food sources. More generally, specialized microbial associations are virtually ubiquitous among animals which, like most herbivores, have diets lacking or imbalanced in the concentrations of important nutrients (Moran *et al.* 2008). In addition to complementation of dietary deficiencies, animals have evolved to take advantage of microbial genes encoding an incredible range of other functions, from camouflaging counter-illumination in squid (Boettcher *et al.* 1996) to anti-parasitoid toxins and modulation of heat tolerance in aphids (Oliver *et al.* 2010).

Increasingly, it is becoming apparent that communities of microbes also affect animal phenotypes in more complex ways. Compared to nutritive symbioses, many of these phenotypic effects are less easily understood as fitness-enhancing expansions of the host's genomic repertoire. The presence of commensal microbes is important to normal development of the gut in *Drosophila* (Shin *et al.* 2011; Broderick *et al.* 2014), but the community shows little association with fly diet or phylogeny (Chandler *et al.* 2011; Wong *et al.* 2013). In mice, obese phenotypes initially caused by knockouts in the nuclear genome can be transferred to wild-type individuals via the microbiome (Turnbaugh *et al.* 2006). Commensal microbial communities have also been shown to affect host reproductive fitness more directly: differences in gut microbiome lead to differences in sexual selection in *Drosophila* (Sharon *et al.* 2010), and immunity-related responses to gut microbial communities may contribute to reproductive isolation in closely related species of parasitoid wasps (Brucker & Bordenstein 2013) and natural populations of house mice (Wang *et al.* 2015).

Mechanisms for the persistence and evolution of host-associated microbiomes

How can selection result in the modification of these phenotypes? Unlike those encoded by the host's own genome, symbiotic phenotypes are not necessarily directly heritable, depending instead on the repeated association of separate genetic replicators with potentially independent demographic histories. Mechanisms to coordinate these interactions have been proposed under a theoretical framework describing the evolution of mutualism between species as falling into three primary types: partner fidelity, partner choice, and byproduct mutualism (Sachs *et al.* 2004; Archetti *et al.* 2011). Under partner fidelity, limitations to symbiont dispersal between host lineages may ensure that subsequent generations of hosts and symbionts interact with one another. Such 'vertical' transmission of the interaction, if sufficiently strict, should link both phylogenetic histories and fitness outcomes for hosts and symbionts, allowing selection on a single symbiotic phenotype to concomitantly influence allele frequency in both partners. In 'horizontal' transmission under partner choice, symbionts could disperse independently of hosts, with reassociation of subsequent generations mediated by some sort of selective mechanism. This might involve specific recognition mechanisms, in which case selection could result in complex, partner-specific phenotypes via a process of reciprocal coevolution. Alternatively, as in the case of byproduct mutualisms, non-specialized partners could be selected from a diverse population of potential partners in each generation based solely on the symbiosis-relevant phenotype itself.

Good examples of both partner fidelity and partner choice mechanisms can be found in nature, particularly among invertebrates with specialized, singular microbial partnerships. Intracellular endosymbionts in insects are frequently transmitted between generations through the egg, the strict partner fidelity resulting in perfectly or near-perfectly linked fitness outcomes and parallel host and symbiont phylogenies across millions of years of evolution (Moran *et al.* 2008). Even in the absence of intracellularity and ovarial transmission, as in the specialized gut symbionts of acanthosomatid stinkbugs, behaviorally-mediated vertical transmission can result in similar outcomes (Kikuchi *et al.* 2009). Yet in other stinkbugs, specialized microbial symbionts appear to be selected from the environment via some kind of partner choice mechanism (Kikuchi *et al.* 2007). One of the best characterized partner choice mechanisms comes from the colonization of light organs in juvenile bobtail squid by their extracellular *Vibrio* symbionts (Nyholm & McFall-Ngai 2004). Many of the known chemoautotrophic endosymbionts of hydrothermal vent invertebrates are similarly selected from the environment by juveniles (Dubilier *et al.* 2008). Compared to the vertically-transmitted endosymbionts of insects, these horizontally-transmitted interactions are less likely to result in symbiont phylogenies that mirror the host phylogeny .

Given the success of these models in describing the highly specific singular associations in invertebrates, they have been attractive candidates for interpreting the more complex communities being revealed by high-throughput sequencing. Evidence supporting both partner fidelity and partner choice mechanisms has been found in more complex gut-associated microbiota. Phylogenies of lower termites, some lineages of their cellulose-degrading protist gut symbionts, and the bacterial ectosymbionts of those protists are all significantly correlated (Noda *et al.* 2007), suggesting that partner fidelity has been maintained over long time scales across multiple levels of association. Vertical transmission of *Helicobacter pylori* has been inferred in humans from the close association between human mitochondrial DNA and *H. pylori* genotype (Falush *et al.* 2003); and although *H. pylori* is typically considered a pathogen, pathology is ameliorated in humans hosting strains matched to their nuclear genotype—consistent with the prediction that partner fidelity aligns fitness outcomes (Kodaman *et al.* 2014). Partner choice mechanisms seem likely to play a more important role in structuring overall vertebrate gut microbiome diversity, as demonstrated by rapid reversion toward host-type communities when mouse and zebrafish gut flora are reciprocally transplanted (Rawls *et al.* 2006). Recent work has begun to uncover some of the molecular mechanisms that may underly partner choice in the vertebrate gut, with complex mucopolysaccharides serving as highly specific selective agents encouraging the colonization of intestinal crypts by particular lineages of *Bacteroides* bacteria (Lee *et al.* 2013).

Over longer evolutionary timescales, we expect these different mechanisms for stabilizing associations to result in differing patterns of diversity among symbiotic microbial communities. For systems exhibiting strong partner fidelity, the shared demographic history of interacting partners should result in correlated phylogenies; the presence or absence of these correlated phylogenetic signals provides a pattern that, while not strictly diagnostic, is at least theoretically testable in the more complex microbial ecosystems found in most animal guts. Systems stabilized by partner choice mechanisms should show similar microbial associates among different individuals within a host species, with the extent of that community similarity across the host phylogeny contingent on the specificity and rate of evolution of the selective mechanisms. Systems that rely simply on the selection by phenotype of microbes from the environment may show very little pattern of association at all, depending on the distribution of relevant phenotypes among environmental bacteria.

Conflicting signals in microbiome diversity

The explosion of data afforded by recent advances in sequencing technology gives us unprecedented opportunity to explore these patterns in natural systems of host-associated microbes. However, interpretation of these data has been hampered by the complexity of the communities being studied. In any given system, different lineages within the microbial community may follow different evolutionary trajectories with respect to the host animal lineage—some components may be largely inherited, some specifically acquired from the environment through coevolved recognition mechanisms, some phenotypically selected or only transiently present – with the resulting patterns of diversity among entire communities thus potentially the result of conflicting signals.

These conflicting signals are perhaps best illustrated by a series of foundational studies describing patterns of diversity in the microbiota of mammals. In the first broad comparative survey of gut microbial diversity across mammalian lineages, Ley and colleagues showed that host diet, not phylogeny, was the strongest predictor of microbiota composition (Ley *et al.* 2008), suggesting that some sort of selective mechanisms were primarily responsible for the observed patterns. The primacy of diet in structuring mammalian microbiota confirmed the evolutionary scope of the selective processes demonstrated in earlier experimental reciprocal transplants of microbiota between host lineages (Rawls *et al.* 2006). A subsequent survey of functional genetic diversity among mammalian microbiomes showed that the same patterns could be recovered from functional gene profiles, demonstrating that taxonomic differences among microbiota corresponded to differences in functional ecology that could conceivably serve as selective environmental filters (Muegge *et al.* 2011).

But despite the prominence of host diet in describing diversity among microbiota, some correlation with host phylogeny could still be observed, exemplified by the overall similarity of the microbiota in carnivorous, omnivorous, and herbivorous lineages of bears (Ley *et al.* 2008). Later described as "phylogenetic inertia" (Delsuc *et al.* 2014), these correlations could conceivably have resulted from the historical conservation of microbial symbionts within the host lineage via a process of long-term partner fidelity resulting in codiversification. Correlation of microbial community with host phylogeny has similarly been interpreted as evidence of long-term vertical inheritance in great ape gut microbiota (Ochman *et al.* 2010), but subsequent investigations have failed to find evidence for this kind of inheritance in more detailed genealogical investigations of wild hominids (Moeller *et al.* 2012; Degnan *et al.* 2012; Moeller *et al.* 2013). While the example of *Helicobacter* codiversification with human maternal lineages (Falush *et al.* 2003) suggests that long-term vertical transmission can occur in the mammalian gut, it is likely that at least some of the observed patterns of correlation between microbiota and host phylogeny are due to other mechanisms.

Topics discussed in this dissertation

Disentangling the relative contributions of these alternative mechanisms will be critical to understanding how association with diverse communities of microbes has affected the course of animal evolution, and how these associations have shaped the evolution of the associated microbial partners. In this dissertation, I will try to tease apart the historical and ecological factors shaping the gut microbiota of animals across millions of years of evolution.

As models, I produce and draw from datasets describing the microbiota of ants and mammals. Both of these animal lineages originated in the Jurassic and experienced dramatic radiations in lineage diversity in the Cretaceous (Moreau *et al.* 2006; Bininda-Emonds *et al.* 2007; Meredith *et al.* 2011). Both mammals and ants have also been shown to host microbial communities corresponding in some degree to diet, especially herbivory (Ley *et al.* 2008; Russell *et al.* 2009; Muegge *et al.* 2011).

In Chapter 1, I explore patterns of phylogenetic correlation in gut microbiota of herbivorous *Cephalotes* ants and hominid apes. Correlation between microbial community similarity and host phylogeny has been used to infer a history of vertical transmission and codiversification of hosts and symbionts (Ochman *et al.* 2010; Rahman 2015), but could also result from horizontal transmission coupled with selective mechanisms that themselves are correlated with host phylogeny. By examining the sensitivity of phylogenetic correlation to analytical parameters, I show that the patterns of similarity between these disparate host lineages and their microbial associates are likely to be result of very different processes.

Next, I examine in more depth the mechanisms likely to underly the interacting effects of diet and phylogeny on the structure of mammalian microbiomes. Like the herbivorous and omnivorous bears nested within a carnivorous clade of mammals, whales consume a diet that differs dramatically from that of their closest extant relatives (Gatesy *et al.* 2013). In Chapter 2, I use a combination of marker gene and shotgun metagenomic sequencing to describe the taxonomic and functional diversity of the whale gut microbiome, providing insight into the specific elements of diet and host physiology most likely to play important roles in driving patterns of microbiome diversity.

Finally, I return to ants to examine how major shifts in the nature of gut microbial association correspond to host ecology. In Chapter 3, I use measures of absolute bacterial abundance, rather than diversity, to test the hypothesis that evolution of symbiosis with microbes has facilitated ants' dominance of tropical rainforest canopies. Surprisingly, I find differences in the abundance of gut bacteria in different ant lineages that span many orders of magnitude, suggesting that evolutionary transitions in the functional role of symbiosis in this animal lineage correspond not only to changes in the diversity of these associations, but to fundamental changes in kind.

Microbes represent one of the greatest potential sources of phenotypic novelty available to animals. Consequently, evolutionary changes that affect an animal lineage's ability to draw from and manage this source of novelty are likely to have profound implications for the subsequent course of evolution. Innovations such as eusociality (which may facilitate partner fidelity through enhanced vertical transmission) and adaptive immunity (which may facilitate partner choice through increased specificity of recognition and sensitivity of response [McFall-Ngai 2007]) could represent fundamental shifts in the macroevolutionary trajectory of animal lineages. Advances in our ability to describe these communities of microbial associates through sequencing have dramatically increased the potential for us to look for evidence of these kinds of transitions, but our ability to analyze and interpret such data remains in its infancy. The data and techniques presented here hint at such fundamental differences in two important animal lineages, and provide the foundation for deeper exploration of the coevolutionary history of animal microbiomes in the future.

CHAPTER 1

Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes[1]

Co-authors: Scott Powell, Daniel J.C. Kronauer, Heraldo L. Vasconcelos, Megan E. Frederickson, Naomi E. Pierce

Abstract

Correlation between gut microbiota and host phylogeny could reflect codiversification over shared evolutionary history, or a selective environment that is more similar in related hosts. These alternatives imply substantial differences in the relationship between host and symbiont, but can they be distinguished based on patterns in the community data themselves?

Here, we explore patterns of phylogenetic correlation in the distribution of gut bacteria among species of turtle ants (genus *Cephalotes*), which host a dense gut microbial community. We use 16S rRNA pyrosequencing from 25 *Cephalotes* species to show that their gut community is remarkably stable, from the colony to the genus level. Despite this overall similarity, the existing differences among species' microbiota significantly correlate with host phylogeny. We introduce a novel analytical technique to test whether these phylogenetic correlations are derived from recent bacterial evolution, as would be expected in the case of codiversification, or from broader shifts more likely to reflect environmental filters imposed by factors like diet or habitat. We also test this technique on a published dataset of ape microbiota, confirming earlier results while revealing previously undescribed patterns of phylogenetic correlation.

---

[1] Chapter published: Sanders, J. G., Powell, S., Kronauer, D. J. C., Vasconcelos, H. L., Frederickson, M. E. and Pierce, N. E. (2014), Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes. Molecular Ecology, 23: 1268–1283. doi: 10.1111/mec.12611

Our results indicate a high degree of partner fidelity in the *Cephalotes* microbiota, suggesting that vertical transmission of the entire community could play an important role in the evolution and maintenance of the association. As additional comparative microbiota data become available, the techniques presented here can be used to explore trends in the evolution of host-associated microbial communities.

Introduction

Gut microbes have had an enormous impact on animal evolution (McFall-Ngai *et al.* 2013). In addition to their well-documented nutritional role in herbivorous mammals, gut bacteria have recently been implicated in processes ranging from brain development in mice (Diaz Heijtz *et al.* 2011) to sexual selection in *Drosophila* (Sharon *et al.* 2010). But our understanding of the mechanisms governing the ecology and evolution of these communities is still hampered by a paucity of comparative data.

In mammals, hosts to perhaps the best-studied animal microbiota, comparative analysis has revealed remarkably consistent correlations between host diet and microbiota composition (Ley *et al.* 2008; Muegge *et al.* 2011), with host lineages sharing convergent gut physiology and diet also tending to share similar gut microbes. Microbiota composition has also been shown to correlate with host phylogeny in the great apes (Ochman *et al.* 2010), perhaps mediated by maternal (i.e., vertical) transmission of microbes, as has been observed in humans for *Helicobacter pylori* (Falush *et al.* 2003). However, subsequent studies of ape microbiota have found the picture to be substantially more complicated than simply one dominated by vertical transmission: maternal lineage explained only a small proportion of variance among chimpanzee microbiota (Degnan *et al.* 2012), while individual chimpanzees switched gut community "enterotypes" over time (Moeller *et al.* 2012). While humans have recently been shown to maintain individual microbial taxa across several years (Faith *et al.* 2013), and to share a somewhat greater proportion of microbial taxa with relatives than with unrelated individuals (Turnbaugh *et al.* 2009, Faith *et al.* 2013), how these dynamics scale across millions of years remains uncertain.

Evidence for consistent trends in the evolution of insect gut microbiota is mixed. Compared to vertebrates, insects tend to have less diverse, and potentially more labile, gut microbial associations (Dillon & Dillon 2004, Engel and Moran 2013). Recent studies have examined gut microbiota of drosphilid flies (Chandler *et al.* 2011), mosquitos (Osei-Poku *et al.* 2012), and leaf beetles (Kelley & Dobler 2010) using next-generation sequencing in a comparative framework; but in each case, microbiota composition has shown little if any correlation with host phylogeny. Where targeted sequencing approaches have revealed extracellular gut symbionts that closely track the evolutionary history of their hosts, it has been in highly specialized cases: stink bugs that sequester symbiont monocultures in gut crypts (Kikuchi *et al.* 2009), and termites whose codiversifying bacteria are themselves physically associated with protists (Noda *et al.* 2007). Detailed comparative surveys of the entire gut microbiota in these organisms are still lacking, limiting our ability to understand how these patterns translate from microbial lineage to microbiota.

For all of these comparative analyses of microbiota, assessing the importance of phylogenetic correlation at the community level is hampered by the potential input of multiple causative factors. Even while some microbial lineages (like *Helicobacter* in humans) are reliably passed vertically from mother to offspring, the bulk of the community may be acquired horizontally from environmental sources. Random horizontal transmission would simply obscure underlying patterns of phylogenetic correlation. But if the ecological success of particular horizontally acquired gut microbes is dependent on selective conditions that are themselves strongly conserved across the host phylogeny (e.g. diet, habitat, or immunity), phylogenetic correlation would be observed even in the absence of vertical transmission.

In this paper, we present one of the first comparative studies of ant gut microbiota using next-generation sequencing, along with a new approach designed to untangle the factors underlying correlation between host phylogeny and microbiota composition. Among insects, ants make a particularly appealing system for studying gut microbiota ecology and evolution. Like mammals, ants utilize a broad range of diets (Davidson *et al.* 2003), permitting comparisons among convergently evolved hosts. Like many other insects, some ant lineages harbor vertically transmitted endosymbionts, which may eventually permit tests comparing the roles of intracellular and extracellular symbioses. We target the arboreal ant genus *Cephalotes*, known as ‘turtle ants’ for their shell-like exoskeletal armor, which host a dense gut microbiota in their midgut and morphologically elaborated hindgut (Fig. A1.1; Roche & Wheeler 1997; Bution & Caetano 2008; Anderson *et al.* 2012).

Although no unequivocal demonstration has been published to date, circumstantial evidence suggests that *Cephalotes* relies upon its specialized gut microbiota to complement a nutritionally imbalanced diet (Jaffe *et al.* 2001). While the details of their feeding biology are still not well understood, *Cephalotes* are, like many other tree-nesting ants, thought to subsist largely on plant-derived nutrients (de Andrade & Baroni Urbani 1999; Davidson *et al.* 2003), including pollen, sap from leaf wounds, insect honeydew, and extrafloral nectar (Byk & Del-Claro 2010; Gordon 2012). This is complemented by feeding on vertebrate waste products, and especially bird droppings (Weber 1957; Adams 1990; Jaffe *et al.* 2001; Powell 2008). Captive colonies fed artificial diets maintain their microbial communities over long timescales (Russell *et al.* 2009, Hu *et al.* 2014), and many of the bacterial lineages appear to be specific to the genus (Anderson *et al.* 2012), giving this association two of the hallmarks of a coevolved mutualistic relationship.

Utilizing 454 sequencing of bacterial 16S genes from 25 *Cephalotes* and several outgroup species, we show that cephalotine ants (*Cephalotes* plus their sister genus *Procryptocerus*) host a relatively simple and remarkably stable microbiota. In contrast both to the non-cephalotine ants in this study, we find that *Cephalotes* nestmates harbor very similar communities. Finally, we use a novel application of sensitivity analysis of microbiota clustering to investigate the role of host phylogeny in explaining the distribution of microbial diversity within the genus *Cephalotes*. This novel analysis approach is further tested and validated using a reanalysis of the phylogenetically-correlated ape microbiota dataset from Ochman et al (2010) for comparison.

Methods

Sample collection and preservation

We collected samples to permit comparisons across the *Cephalotes* host phylogeny, and also to explore the influence of geography and colony structure on microbiota composition. To that end, we concentrated novel collections at two field sites with high *Cephalotes* species diversity: the Panga Ecological Station (Estação Ecológica do Panga) in Minas Gerais, Brazil (19°10' S, 48°23' W); and the Los Amigos research station (Centro de Investigación y Capacitación de Rio Los Amigos) in Madre de Dios, Peru (12° 33' S, 70° 8' W). These two sites represent very different ecosystems, contributing both to the potential phylogenetic diversity of host species and to ecological diversity with respect to exogenous microbes. The Brazil site is a dry neotropical savannah biome, or 'cerrado,' with 3-5m tall trees providing 30-50% canopy cover. The Peru site is dominated by primary wet tropical forest, approaching 100% canopy cover at a canopy height of ~30m. In total, we were able to include 25 *Cephalotes* colonies, representing 17 unique species, from these two locations (for detailed collections information, see Table A1.1). Four species (*C. minutus, C. atratus, C. maculatus,* and *C. clypeatus*) are represented in the dataset by a colony collected in each location. To permit additional intra-specific, inter-colony comparisons, two abundant species from the Brazil site (*C. pusillus* and *C. persimilis*) are represented by three colonies each. Three individuals from *C. pusillus* colony #12 were included on two separate sequencing rounds, serving as example technical replicates. Most colonies were discovered by baiting trees with nitrogen-rich baits (as in Powell 2008), with some additional colonies discovered by twig snapping.

This primary sample set was augmented with additional samples to provide broader coverage of the *Cephalotes* phylogeny as well as outgroups for genus-level comparisons (Table A1.1). We collected one additional *Cephalotes* colony (*C. rohweri*) from desert scrubland in Arizona, and included individual ants from 7 additional *Cephalotes* species from ethanol-preserved museum collections. For outgroups, we included 2 ethanol-preserved individuals from separate museum collections of *Procryptocerus* (the sister genus to *Cephalotes),* as well as one colony from each of the arboreal ants *Crematogaster*, *Azteca*, and *Pseudomyrmex* collected from the Brazil field site concurrently with *Cephalotes* colonies.

We processed samples to maximize the representation of gut microbes while trying to limit the influence of exogenous contamination from either environmental or host-derived, non-gut microbes. Thus, for the colonies collected specifically for this study (excepting museum samples and workers from the *C. rohweri* colony, which were kept for several weeks in the laboratory before being dissected and immediately extracted without intermediate preservation), guts were dissected in the field within 2 days of collection. Adult workers were killed in 100% ethanol, transferred within 30 minutes to a 1:10 solution of bleach in distilled water (final concentration ~0.5% sodium hypochlorite) for

30-60 seconds, and rinsed in filter-sterilized phosphate-buffered saline (PBS). Dissections took place in sterile PBS. Dissected guts were transferred immediately to filter-sterilized RNALater nucleic acid preservative (Ambion, Inc). To maximize consistency among samples, only the midgut and ileum (excluding the fragile crop and rectum) were extracted for sequencing. For freshly collected specimens, three workers were sequenced separately per colony.

Since ethanol dehydration renders internal structures extremely fragile, museum specimens were not dissected. Instead, whole individual ants were transferred from ethanol to bleach solution as above, rinsed in PBS, and just the gasters retained for analysis. Due to limited specimen availability, only a single worker was sequenced per colony for these species. To help evaluate the potential biases imposed by alternative preservation methods, three additional workers from one of the freshly-collected *C. pusillus* colonies were preserved in ethanol and processed in the same manner as the museum specimens.

DNA extraction and sequencing

We employed a relatively intensive extraction protocol, based on a method developed for sampling termite gut microbes (Matson *et al.* 2007), to minimize the potential for biases against difficult-to-lyse microbes (e.g. Firmicutes, see Willner *et al.* 2012) and to decrease the potential influence of inhibitors on downstream enzymatic reactions. Briefly, tubes containing preserved guts in RNAlater were diluted ~1:1 with sterile water (to decrease solution density and dissolve any precipitated salts) and spun for 10 minutes at 14,000 rcf. Supernatant was removed, and replaced with 700μL of TLS-C sample lysis buffer (MPBio, inc). Tubes were then vortexed at maximum speed for 1 minute to resuspend pelleted material. Contents were transferred to sterile lysis tubes containing a bead mixture (Lysis matrix A, MPBio) and 500μL Phenol:Chloroform:Iso-

Amyl Alcohol, pH 8, and mechanically lysed for 40 seconds at maximum speed on an MPBio FastPrep-20. After lysis, tubes were centrifuged at 8,000 rcf for 1 minute and the aqueous phase removed and washed with 500µL chloroform. The remaining aqueous phase was column purified using Qiagen DNeasy Blood and Tissue extraction columns, starting with addition of equal volumes of buffer AL and 100% molecular grade ethanol to the aqueous phase and application to the column. The remainder of the cleanup was performed according to manufacturer's instructions. Finally, purified extracts were concentrated by isopropanol precipitation, resuspended in 32µL TE, and quantified using a Qubit fluorometer (Invitrogen).

To characterize the microbial community, an approximately 500bp fragment spanning the V1-V3 regions of the bacterial 16S gene was amplified with universal primers 27F and 515R (see Kumar *et al.* 2011) and sequenced using 454 Titanium chemistry at a commercial facility (Research and Testing Laboratories, Lubbock, TX) according to previously published protocols (Dowd *et al.* 2008). Briefly, amplifications were performed in 25 ul reactions using 1ul of each 5uM primer and 1ul of template. Reactions were performed under the following thermal profile: 95°C for 5 min, then 35 cycles of 94°C for 30 sec, 54°C for 40 sec, 72°C for 1 min, followed by one cycle of 72°C for 10 min.

16S rRNA sequence filtering and clustering

All microbiota community sequences were processed to limit the effect of amplification and sequencing artifacts using AmpliconNoise v1.25 under parameters recommended for 454 Titanium chemistry (Quince *et al.* 2009; see Appendix I for a brief discussion of this process). Denoised sequences were then analyzed using QIIME v1.4.0 (Caporaso *et al.* 2010). Some samples initially yielded fewer sequences than we specified, and the same amplicon pools were re-run by the sequencing facility; in these cases, we excluded sequences from the poorer initial run from further analysis. Specific analysis scripts are detailed in Appendix I, but with the exception of modifications to sequence clustering detailed below, most steps were performed using default parameters. We assigned taxonomy to sequences with the RDP classifier (Wang *et al.* 2007) using the curated GreenGenes 16S rRNA database dated 4 February 2011, available from the QIIME website (Werner *et al.* 2011).

We clustered sequences into Operational Taxonomic Units (OTUs) using two clustering algorithms and a total of four different similarity thresholds (93, 95, 97, and 99%). Chimeric sequences were removed using both *de novo* and reference-based chimera removal (using the Gold 16S database described in Haas *et al.* 2011) with UCHIME (Edgar *et al.* 2011). Clustering sequences into OTUs reduces the impact of sequencing error, speeds computation, and permits analyses unbiased by assumptions about bacterial taxonomy. However, different OTU clustering algorithms may give very different -- and sometimes overstated -- estimates of bacterial diversity in a sample (Huse *et al.* 2010). Furthermore, while different percent similarity thresholds should theoretically correspond to different average evolutionary divergence times, various clustering algorithms can yield surprisingly different results at a given similarity threshold (Sun *et al.* 2012). Thus, we performed separate analyses at 93, 95, 97, and 99% similarity thresholds using the CD-HIT (Li & Godzik 2006) and UCLUST (Edgar 2010) algorithms implemented in QIIME. Both algorithms were run under default parameters.

Comparative analyses

We calculated basic descriptive and comparative statistics for microbiota data under all clustering parameter combinations using QIIME 1.4.0. To permit summarized comparisons across the broadest possible range of samples, alpha diversity estimates (including observed species richness, Shannon diversity, the Chao1 nonparametric richness estimator, and whole-tree phylogenetic diversity) were calculated using sample by OTU abundance observation tables (OTU tables) rarified to 1000 observations per sample and excluding samples with less than 1000 high-quality sequences. Beta diversity metrics (including abundance weighted and unweighted UniFrac distances and Sørensen, Jaccard, and Bray-Curtis dissimilarities) were calculated using OTU tables rarified to 1000 observations, but retaining samples with fewer sequences (see Aguirre de Cárcer *et al.* 2011). Alpha and beta diversity calculations, including collectors' curves, PCoA calculations, and OTU network tables, were generated in QIIME. OTU networks were visualized in Cytoscape v2.8.1 (Shannon *et al.* 2003). We also calculated the 'core' OTUs (present in ≥ 50% of *Cephalotes* samples) for 93, 95, and 97% thresholds using QIIME v1.6.0.

We used a number of methods to explore the impacts of colony structure, host phylogeny, and geography on microbiota composition. Between-sample geographic distances were calculated from sample locality information using the AMNH geographic distance calculator tool (Ersts 2013). Host genetic distances were calculated as patristic distances in PyCogent (Knight *et al.* 2007) using the time-calibrated *Cephalotes* phylogeny of Price *et al.* 2014, modified by hand to include the outgroup genera *Pseudomyrmex*, *Azteca*, and *Crematogaster* with approximate branching times as indicated in the phylogeny of the ants by Moreau *et al.* (2006).

Beta diversity dissimilarity matrices, along with host geographic and genetic distances, were imported into R (R Development Core Team 2012) for further analysis. To test for differences in average between-sample distances among sample categories (e.g. across levels of host colony, species, clade [*sensu* de Andrade & Baroni Urbani 1999] and genus), we used Monte Carlo permutations of category labels to generate null distributions of between-sample distances appropriate to each comparison; direct comparisons of pairwise distances violate assumptions of independence for most other statistical tests. Specific permutation designs are described in appendix I. To test for significant associations between bacterial community dissimilarities and host genetic and geographic distances, we used partial Mantel tests, as implemented in the vegan package in R (Oksanen *et al.* 2012).

Beta diversity clustering sensitivity analysis

Clustering of microbial communities by similarity has been shown to recapitulate host phylogeny in a few cases, such as between species of great apes and their microbiota (Ochman *et al.* 2010), lending support to the idea that gut communities coevolve with their hosts. However, sample clustering is strongly influenced by OTU picking parameters and choice of diversity metric (Hamady & Knight 2009). Changes in the pattern of support under these different parameter conditions may be useful in interpreting biological significance. For microbial communities that correlate with host phylogeny, we would expect measures of beta-diversity to be differentially affected by changes to OTU clustering, depending on how these correlations arose (see Fig. 1.1). For example, if neutral codiversification were the sole force shaping the gut community of a 15 million-

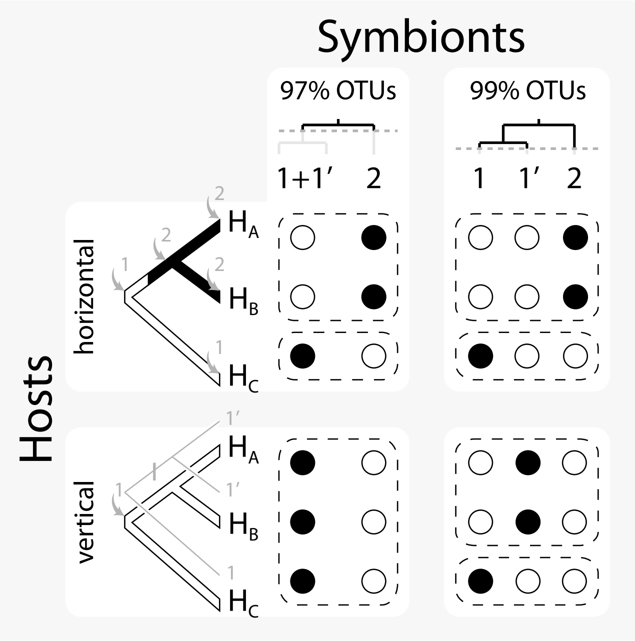


Figure 1.1: Conceptual illustration of the beta diversity sensitivity analysis. Host ($H_{A,B,C}$) phylogenies are depicted on the left, symbiont (1, 1', 2) phylogenies above. In each scenario, presence/absence of symbionts in each host is depicted by filled/open circles, respectively. Similar host communities for each scenario are grouped by a dotted border. Top row: hosts select symbionts horizontally from the environment with each generation. A change in host diet (depicted by black fill on host tree) leads to the replacement of symbiont 1 by the symbiont 2 in H_A and H_B. Since symbionts 1 and 2 diverged long ago, they fall into different OTUs at both 97% and 99% clustering widths, so grouping of microbiota H_A and H_B separately from H_c is insensitive to clustering width. Bottom row: hosts acquire symbionts vertically from parent generation. A mutation in symbiont 1' (depicted by a vertical hash) causes sequence divergence sufficient to cluster separately from symbiont 1 under 99% OTU clustering, but not under 97%. Consequently, grouping of microbiota H_A and H_B separately from H_c is sensitive to clustering width.

year-old host genus, and mutations in 16S rRNA accumulated uniformly among gut microbes at a rate of 0.1% per million years, homologous gut microbes -- meaning, specifically, those derived from a shared ancestral microbe -- across the genus should be at least 98.5% identical (or at most 1.5% different) at the 16S locus. In this case, clustering microbial OTUs at 97% similarity or below would cause all hosts in the genus to appear to host identical gut communities, and microbiota from closely related hosts would not be expected to appear more similar. Clustering OTUs at 99% would start to reveal phylogenetically-correlated microbiota structure, with host clades separated by less than 10my grouping together. Thus, the constraint on genetic distances between microbes imposed by the age of the hosts' most recent common ancestor would be reflected in a threshold OTU clustering width, below which beta-diversity metrics would no longer be reflective of host phylogeny. By contrast, when correlation of microbiota structure with host phylogeny is mediated by environmental factors (e.g. differences in diet that in turn correlate with host phylogeny), we would expect beta diversity metrics to reflect host phylogeny with much less sensitivity to OTU clustering width.

We would also expect different measures of beta diversity to perform differently in these cases. "Star phylogeny" measures of beta diversity, such as Sorensen and Jaccard dissimilarities, weight each OTU equally, regardless of how closely related two different OTUs might be (Lozupone & Knight 2008). These methods effectively increase the sensitivity of the metric to recent bacterial evolution, since divergence just exceeding the OTU clustering threshold will have the same effect as much older splits. By contrast, the UniFrac metric is designed to minimize the effects of such recent bacterial evolution by weighting the longer, internal branches of the bacterial phylogeny. Thus, we would expect communities differing primarily due to recent diversification to separate more clearly using Jaccard dissimilarities than UniFrac distances.

Borrowing a technique from systematics (Sanders 2010), we visualized the sensitivity of beta-diversity-based sample clustering to various parameter combinations using a series of grids overlaid on the host phylogeny. We generated UPGMA-clustered dendrograms of 100 jackknifed OTU tables for all five beta-diversity metrics and four OTU clustering thresholds, and compared these to the host phylogeny using the tree_compare.py script in QIIME. We performed this analysis for our entire ant dataset, using OTU tables summarized by colony. For context, we also performed this analysis on the great apes microbiota dataset of Ochman *et al.* (2010) mentioned above. Because the ape data were sequenced at greater depth, we repeated the analysis at both a level of rarefaction close to that used for the *Cephalotes* data (1,000 sequences/sample) and at much higher coverage (15,000 sequences /sample) to approximate previously published analyses of these data. Partial Mantel correlations were performed as for the ants, using patristic distances from the time-calibrated whole-genome phylogeny of Prado-Martinez and Sudmant *et al.* (2013) and geographic distances estimated from Figure 1 in Ochman *et al.* (2010) using Google Earth.

Results

Sequencing results and taxonomic composition

454 sequencing reveals that *Cephalotes* host a relatively simple microbiota that is remarkably conserved. After denoising, clustering, and chimera-checking, we generated a total of 241,519 sequences from 102 specimens in a nested design, permitting comparisons within colonies, among conspecific colonies, among *Cephalotes* species, among geographic areas, and between *Cephalotes* and four other genera (see Table A1.2 for per-sample sequence counts).

Cephalotes gut microbiota from across the genus were dominated by Verrucomicrobia and Proteobacteria (Fig. 1.2). This is the first genus-wide survey with broad phylogenetic sampling for *Cephalotes*, and is consistent with previous results from small numbers of species (*C. atratus*, *C. rohweri,* and *C. varians; see* Russell *et al.* 2009; Anderson *et al.* 2012; Kautz *et al.* 2012). These communities appear to be relatively simple, averaging just 20 unique 97% OTUs per 1000 sequences (Fig. A1.2, Table A1.3). Many of these OTUs were widely distributed across the genus, occurring in more than 50% of samples. All of these ‘core OTUs’ were close matches to sequences from clades that have previously been described as *Cephalotes*-specific (Fig. A1.3; Anderson *et al.* 2012). Verrucomicrobia sequences dominated most *Cephalotes* gut samples, though the combined effects of tissue choice and preservative had strong effects on relative abundance (Fig. 1.2; see Appendix Information for additional discussion). The two *Procryptocerus* samples were broadly similar to *Cephalotes* in both taxonomic composition and measures of species richness and alpha diversity (Fig. 1.2, A1.2). By contrast, gut microbiota from non-cephalotines showed very little taxonomic overlap with those from *Cephalotes* (Fig. 1.2), and at least for *Azteca* and *Pseudomyrmex*, were substantially more diverse (Fig. A1.2, Table A1.3). Additional results, including notable trends in relative abundance for particular microbial taxa, can be found in Appendix I.

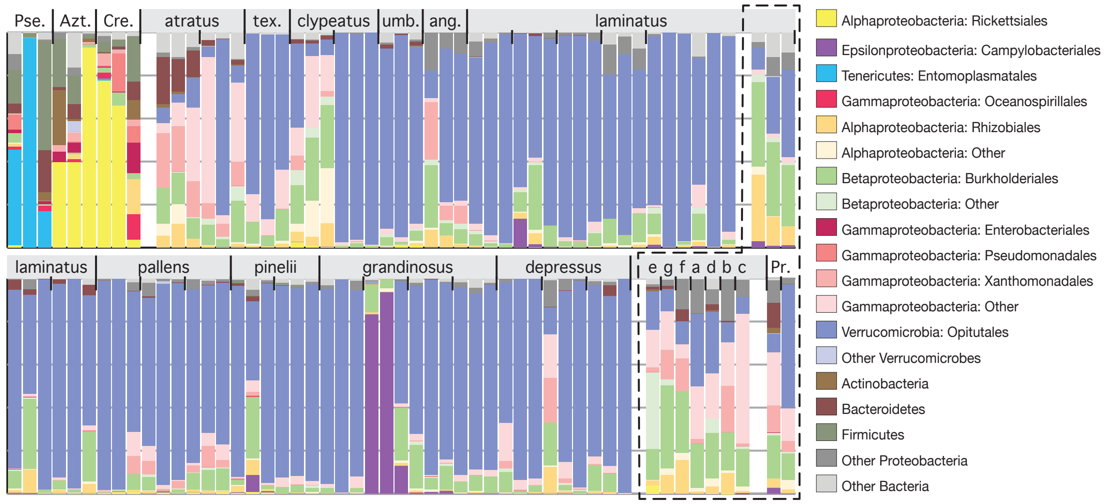


Figure 1.2: Distribution of class-level taxonomic diversity across sampled ant gut microbiota. Bars are placed roughly according to host phylogeny, with small ticks between colonies and large ticks separating host genera/clades. Museum specimens are labeled with colony letter (see Table A1.1 for reference). Blank spaces are left purely to assist in visual separation of groups. Museum samples derived from ethanol-preserved gasters are circled by a dashed line to emphasize the apparent effect on relative abundance of Verrucomicrobia sequences. Three freshly-collected individuals were also preserved in ethanol and sequenced from whole gasters; these are placed on the top right, next to dissected gut-derived samples from the same nest.

Beta diversity: effects of colony structure

Cephalotes nestmates had gut communities considerably more similar to one another than to the gut communities of conspecifics from other colonies or to communities from other species. At the 97% OTU clustering threshold, the average Jaccard dissimilarity among *Cephalotes* nestmates was 0.66 (sd = 0.16, n = 90), indicating that roughly one third of OTUs were shared between individuals from the same colony (Fig. 1.3). This probably overestimates the divergence between samples, as the three technical replicates in our sample showed a comparable level of dissimilarity (mean of 0.74, sd = 0.15, n = 3). Several factors are likely to have contributed to the apparently high technical variance component, perhaps dominated by the relatively high PCR cycle number performed by the sequencing facility (see Appendix I for additional discussion). Given that caveat, nestmates appear to share quite similar communities. By comparison, just 21% of OTUs (Jaccard dissimilarity = 0.79) were shared between conspecific individuals from separate colonies (sd = 0.078, n = 108), only marginally more than the 14% shared between heterospecific *Cephalotes* (sd = 0.072, n = 270; *p* = 0.012). UniFrac distances between samples showed a similar pattern (Fig. A1.2).

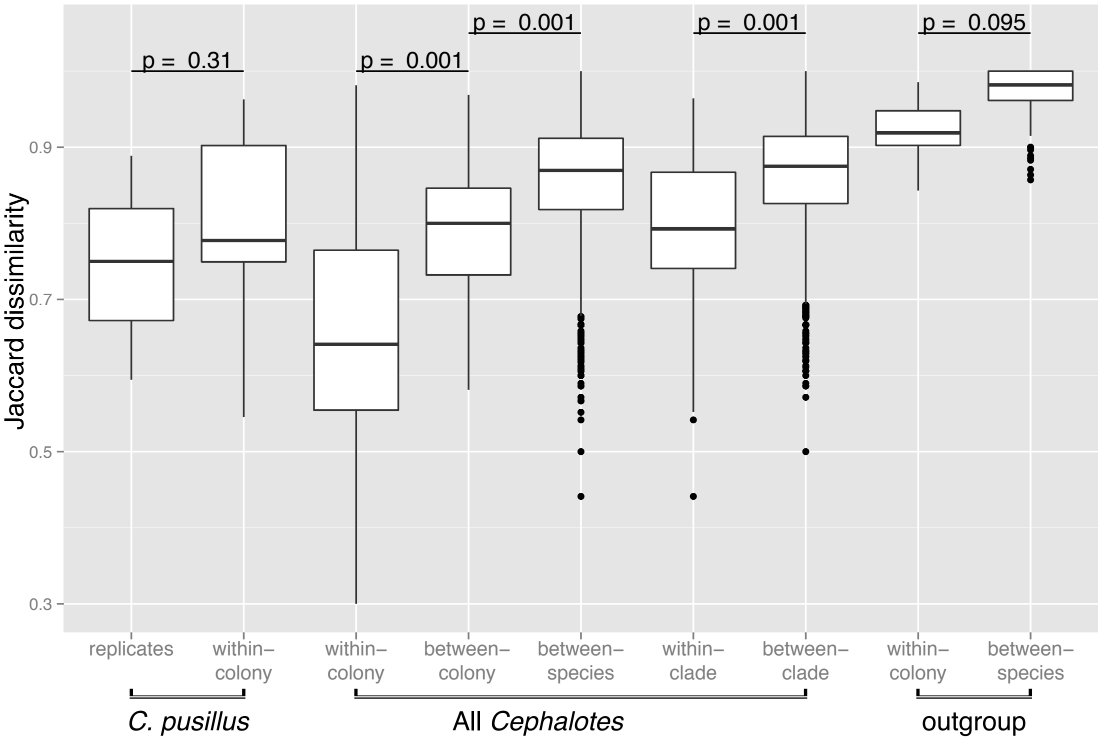


Figure 1.3: Dissimilarity boxplots, comparing community Jaccard dissimilarities within and among groups of ants for 97% OTUs. Significance values for between-group comparisons were calculated using Monte Carlo permutation tests (see Appendix I for details on permutation structure).

In contrast to the consistency observed among *Cephalotes* nestmates, just 8% of OTUs were shared between nestmates of the outgroup genera *Pseudomyrmex*, *Crematogaster*, and *Azteca* (Jaccard dissimilarity = 0.92, sd = 0.030, n = 9). Dissimilarities between heterospecifics among these three genera were higher than within-colony dissimilarities (mean = 0.98, sd = 0.030, n = 93), though not significantly so ($p = 0.095$).

Beta diversity: effects of host phylogeny and biogeography

Network visualization of shared OTUs (Fig. 1.4) suggested that host phylogeny plays an important role in structuring the *Cephalotes* gut community. In the network analysis of 99% OTUs, samples were positioned using a spring-embedded edge-weighted algorithm, which places individual samples closer together solely based on the number of shared OTUs. By coloring sample nodes according to their position on the host phylogeny, clear divisions were visible among *Cephalotes* clades. Divisions between the cephalotine ants and the outgroups, and between *Cephalotes* and *Procryptocerus,* were even more apparent at the wider 97% OTU clustering width (Fig. A1.5).

Principal coordinates analysis of beta diversity dissimilarities (Fig. A1.6-A1.9) largely recapitulated the network analysis, with *Cephalotes* samples from the same clade tending to group together. This effect was particularly apparent with narrower OTU picking thresholds and beta diversity metrics, like Jaccard dissimilarity, dependent solely on the number of shared OTUs (Fig. A1.6). Unweighted UniFrac, which takes into account the phylogenetic similarity of shared OTUs, clearly separated cephalotine samples from outgroups, but was less likely to group samples from related host species, especially among the more recently diverged clades. In general, the separation between the earlier branching groups (especially *C. atratus*) and the remainder of the *Cephalotes* phylogeny was apparent across a broader range of clustering widths and diversity measures. Neither geography (Fig. A1.7), preservation method (Fig. A1.8), nor sequencing quadrant (Fig.

A1.9) appeared to have strong effects relative to host phylogeny with these metrics. Most variance in abundance-weighted metrics, such as Bray-Curtis and Weighted UniFrac, appeared to be driven primarily by differences in the relative abundance of Verrucomicrobia, which in turn was strongly affected by tissue and/or preservative (see Appendix I).

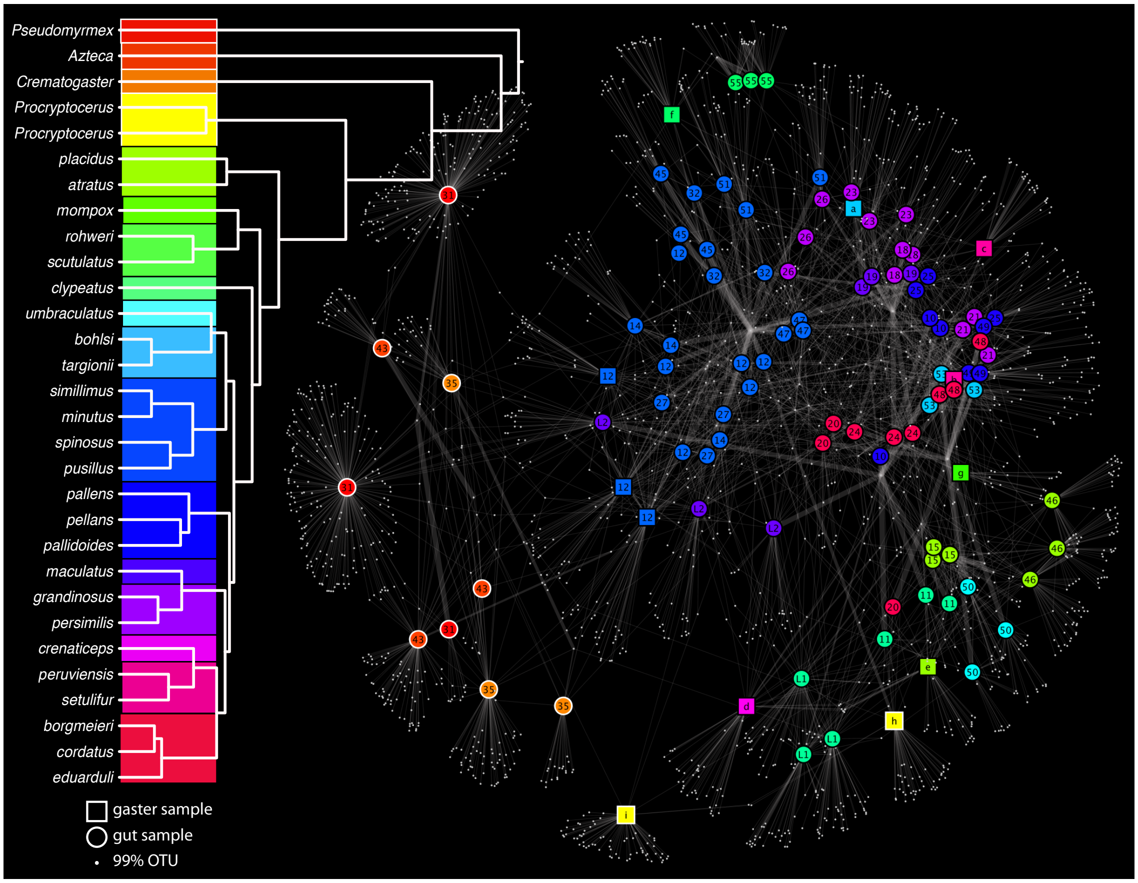

Pseudomyrmex
Azteca
Crematogaster
Procryptocerus
Procryptocerus
placidus
atratus
mompox
rohweri
scutulatus
clypeatus
umbraculatus
bohlsi
targionii
simillimus
minutus
spinosus
pusillus
pallens
pellans
pallidoides
maculatus
grandinosus
persimilis
crenaticeps
peruviensis
setulifur
borgmeieri
cordatus
eduarduli
gaster sample
gut sample
99% OTU

Figure 1.4: OTU network showing relationships among ant gut microbiota. Large, colored nodes represent individual ant samples, while small grey nodes represent individual 99% OTUs. Edges connect OTUs with each host sample in which they occur. Nodes are placed according to a weighted, spring-embedded algorithm, causing host nodes that share more OTUs to appear close together. Host nodes are labeled by colony (see Table A1.1) and colored by clades (*sensu* De Andrade and Baroni-Urbani 1999) in rainbow order according to phylogeny (inset), with non-*Cephalotes* host nodes outlined in white for emphasis.

Partial Mantel tests indicated that the majority of the variance in community beta diversity could be explained by host genetic distance, both for the dataset as a whole, and for the subset of *Cephalotes* samples (Fig. 1.5c). For the whole dataset, after accounting for geographic distance, correlation between host genetic distance and Jaccard community dissimilarity increased at wider OTU clustering thresholds, and ultimately accounted for as much as 79% of variance using 93% OTUs (Fig. 1.5c; correlation between UniFrac community dissimilarity and host genetic distance was insensitive to OTU clustering width [Fig. A1.10]). For comparisons among *Cephalotes*, however, narrower OTU clustering thresholds explained a greater proportion of the total variance, suggesting that much of the among-*Cephalotes* phylogenetically-correlated variance is a consequence of recent bacterial evolution. Genetic distance and community dissimilarity were highly significantly correlated at every OTU clustering threshold for both the whole and *Cephalotes* datasets ($p < 0.001$). Geographic distance also explained a small proportion of the community dissimilarity after correcting for genetic distance, with generally higher correlation at lower OTU clustering thresholds (Fig. 1.5c). Results for the ape microbiota data were substantially different (Fig. 1.5d), showing a much lower overall impact of phylogeny, and only marginal significance at all OTU clustering widths.

Beta diversity clustering sensitivity

To summarize the expectations detailed above, we predicted that communities whose patterns of similarity arose primarily through recent bacterial evolution would be grouped more often at narrower OTU clustering widths than wide ones (Fig. 1.1), and more often using “star phylogeny” measures of beta diversity, such as Jaccard or Sorensen dissimilarities, than using UniFrac.

Our sensitivity analysis approach recovered just such a pattern for the internal nodes of the great ape dataset from Ochman *et al.* (2010), recapitulating the results of that study (Fig. 1.5b). As shown previously, the nodes grouping chimpanzees, chimpanzees + bonobos, and chimpanzees + bonobos + humans were all recovered primarily using 99% OTU clustering and star phylogeny diversity measures, and only at the deeper level of rarefaction. Our analysis also reveals patterns that were not apparent in the earlier, parsimony-based analysis. Strikingly, most of the nodes grouping samples at the tips of the tree -- i.e., from the same species or subspecies -- were recovered robustly under most parameter combinations. This suggests the presence of some common selective filter, such as diet, generating cohesion in these communities (for example, see Fig. 1.1).

Our analysis of *Cephalotes* microbiota showed a similar pattern to that observed for the internal nodes of the ape phylogeny (Fig. 1.5a), suggesting that recent evolution of gut bacteria may mirror host evolution. Using Jaccard dissimilarities, 15 of 28 internal nodes on the *Cephalotes* phylogeny were recovered in at least one jackknife replicate. Of these, 9 nodes were recovered more often under 99% OTU clustering than under any other threshold. In contrast to the ape data, nodes grouping conspecifics were generally not broadly supported, implying that species-specific selective filters don't play as large a role in differentiating *Cephalotes* microbiota from each other. Two notable exceptions to this were the node grouping the two *C. atratus* colonies and that grouping *C. pallidoides* and *C. pellans* (Fig. 1.5a), both of which were supported more often under wider OTU thresholds -- suggesting that similarity among these microbiota may reflect broader shifts in the communities. The nodes separating *Cephalotes* from *Procryptocerus*, and separating the cephalotines from the outgroup genera, were supported at higher frequency and across more measures of beta diversity, reflective of the greater differences between these communities.

To ensure that these results were not simply due to chance, we also compared 99% Jaccard support among clustered *Cephalotes* microbiota to 100 host trees with randomly permuted tips. No nodes were supported under these parameters in 56% of these randomized datasets, with a maximum of three supported in two of the permutations (Fig. A1.10).

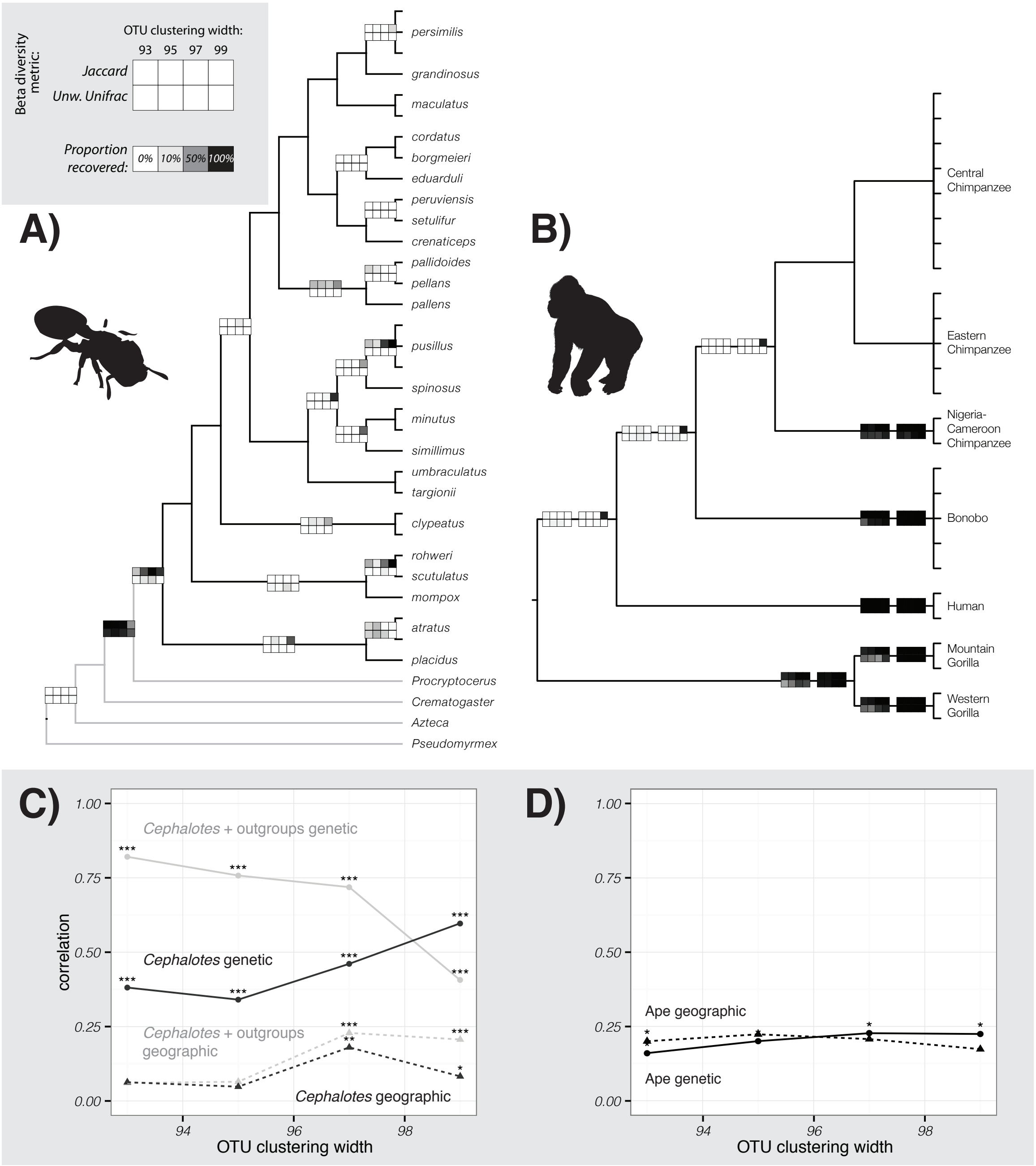

OTU clustering width:
93 95 97 99
Beta diversity metric:
Jaccard
Unw. Unifrac
Proportion recovered:
0%
10%
50%
100%
A)
persimilis
grandinosus
maculatus
cordatus
borgmeieri
eduarduli
peruviensis
setulifur
crenaticeps
pallidoides
pellans
pallens
pusillus
spinosus
minutus
simillimus
umbraculatus
targionii
clypeatus
rohweri
scutulatus
mompox
atratus
placidus
Procryptocerus
Crematogaster
Azteca
Pseudomyrmex
B)
Central Chimpanzee
Eastern Chimpanzee
Nigeria-Cameroon Chimpanzee
Bonobo
Human
Mountain Gorilla
Western Gorilla
C)
correlation
1.00
0.75
0.50
0.25
0.00
Cephalotes + outgroups genetic
Cephalotes genetic
Cephalotes + outgroups geographic
Cephalotes geographic
***
**
*
94
96
98
OTU clustering width
D)
1.00
0.75
0.50
0.25
0.00
Ape geographic
Ape genetic
*
94
96
98
OTU clustering width

Figure 1.5: Contrasting patterns of phylogenetically-correlated microbiota in ants and apes. A and B): beta diversity clustering sensitivity analysis of ant (A) and ape (B) samples using the Jaccard and UniFrac dissimilarity indices (other indices omitted for clarity; for these results, see Fig. A1.11). Each grid represents the support for a particular sample grouping across a jackknifed dataset, with each square representing a continuous gradient of support for that grouping at a given combination of parameters; black = 100%, white = 0%. Parameter combinations are given in the inset key at left. For ape data, grids are displayed for two levels of sampling depth, with results for 1000-sequence rarefactions at left and 15,000 sequence rarefactions at right. Only nodes recovered at least once have grids displayed. For ant data, terminals represent individual colonies; *C. persimilis*, *C. maculatus*, *C. pusillus*, *C. minutus*, *C. clypeatus*, and *C. atratus* are all represented by multiple colonies. C and D) Results from partial Mantel tests for ants (C) and apes (D). Correlation coefficients between Jaccard dissimilarity matrices and genetic and geographic distances, respectively, are plotted at four different OTU clustering thresholds. P-values from partial Mantel tests indicated by asterisks (***, $p = 0.001$; **, $0.001 < p < 0.01$; *, $0.01 < p < 0.05$). Ant microbiota (C) show a strong influence of phylogeny in explaining variation. Results plotted for *Cephalotes*-only and for *Cephalotes* plus outgroups (grey lines). Ape microbiota (D) show less correlation with phylogeny across all OTU clustering widths.

Discussion

Correlation between phylogeny and the gut microbiota

Our results show that the gut microbiota of *Cephalotes* ants are very stable -- perhaps exceptionally so among ants, as even individuals from different *Cephalotes* species typically shared microbiota more similar to one another than did nestmates of the other ant genera we tested (Fig. 1.3). Our findings are the first to characterize gut microbiota across a broad fraction of this genus, one of the most diverse ant lineages in the neotropics (Fernández & Sandoya 2004).

Remarkably, despite this overall similarity among *Cephalotes* gut communities, we observed a substantial and significant effect of host phylogeny. This effect was readily apparent in the clustering of closely related species in OTU network diagrams (Fig. 1.4, A1.5) and PCoA ordinations (Fig. A1.6). Under some parameter sets, host phylogenetic distance accounted for the majority of total variation in beta diversity among *Cephalotes* microbiota (Fig. 1.5c). By comparison, geographic distance accounted for far less of this variation, though our samples were not collected for biogeographic comparisons and are thus limited in this regard.

That closely related host species harbor similar microbiota suggests, but does not necessarily demonstrate, some degree of codiversification between hosts and microbes. In mammals, for example, gut microbes are highly correlated with diet (Ley *et al.* 2008). Similar microbes tend to inhabit the guts of unrelated hosts with convergently evolved diets, indicating that host switching occurs relatively frequently compared to the rate of microbial diversification; thus, very few microbial lineages appear to be restricted to monophyletic groups of mammals (Muegge *et al.* 2011, Delsuc *et al. in press*). In cases where changes in diet or other environmental filters are phylogenetically correlated rather than convergent, clear patterns of correlation between microbiota and host phylogeny could thus be due more to these filters than to a history of codiversification with the host.

We developed our beta diversity clustering sensitivity analysis as a way to help distinguish between these alternative mechanisms -- phylogenetically-correlated environmental filtering and shared evolutionary history -- in explaining correlation between microbial community composition and host phylogeny. The intuition for this approach is based on the assumption that, in either scenario, more distantly related hosts will have more distantly related microbes; but in the case where historical codiversification is the sole or primary factor leading to similarity among microbiota, the age of the last common ancestor of the hosts will constrain the genetic distance between the symbionts. In other words, recent host speciation should be reflected by recent symbiont speciation. By contrast, in the case where host diet selects for different microbes, the most recent common ancestor of a pair of microbes in the two hosts may far pre-date the last common ancestor of the hosts. As a hypothetical example, a particular lineage of bacteria may have diversified into herbivore-gut and carnivore-gut specialist lineages along with the evolution of herbivory in terrestrial vertebrates in the Carboniferous (Sues & Reisz 1998), far predating the evolution of dietary specialization among placental mammals in the Cretaceous (Bininda-Emonds *et al.* 2007).

To test this approach, we reanalyzed the ape microbiota dataset from Ochman *et al.* (2010), a frequently-cited example of codiversification in microbiota. Confirming its utility, our method recapitulated the earlier findings: most internal nodes of the ape phylogeny were recovered under 99% OTU clustering with Jaccard dissimilarity (Fig. 1.5b). By testing additional parameter combinations, though, our method goes further: that these internal nodes were *not* recovered under wider OTU clustering thresholds implies that the information grouping different ape species and subspecies is primarily a product of recent bacterial evolution. This was not previously known, as the earlier analysis only reported results from 99.5% OTUs (Ochman *et al.* 2010). That we were only able to recover support for these internal nodes at greater sequencing depth also suggests that the microbial taxa supporting these groupings make up a relatively small proportion of the community.

Perhaps more surprisingly, the additional parameter combinations in our analysis revealed that the grouping of gut communities from conspecific hosts was highly *insensitive* to parameter choice -- the information grouping communities within subspecies (and, for gorillas, between subspecies as well) is retained even when obscuring many millions of years of bacterial divergence in very broad OTUs, and at relatively shallow depth of sampling. This pattern of broad support suggests an important role for horizontal acquisition of microbiota through species-specific filters, such as diet or immune selectivity. The partial Mantel correlations (Fig. 1.5d), which show that phylogenetic distance explains only a small and marginally significant proportion of ape microbiota beta diversity, echoes this pattern.

Taken together, our results would be consistent with a model in which apes acquire species-specific microbiota largely horizontally, while retaining a small proportion of vertically-transmitted microbes over longer timescales. Our findings may help to reconcile the apparent patterns of codiversification found by Ochman *et al.* (2010) with subsequent studies that found larger roles for factors such as social group affiliation and geography (Degnan *et al.* 2012, Moeller *et al.* 2012, Moeller *et al.* 2013). Notably, we were able to detect both patterns using only the original dataset.

As observed for the internal nodes of the ape phylogeny, our sensitivity analysis suggests that phylogenetic correlation in the *Cephalotes* microbiota is driven in large part by recent bacterial evolution. The results of the partial Mantel tests (Fig. 1.5c) reinforce this finding. Within *Cephalotes*, microbiota at 99% OTU clustering showed much greater correlation with host phylogeny than at wider clustering thresholds. Furthermore, the much greater overall proportion of beta diversity variance explained by the *Cephalotes* phylogeny relative to that explained by the ape phylogeny suggests that such recent evolution (potentially, the result of codiversification) plays an overall greater role in structuring the *Cephalotes* microbiota than it does in the apes'. This interpretation is further supported by the fact that we didn't see a shift to broad, parameter-insensitive support -- possibly indicative of some sort of environmental filtering -- at the tips of the *Cephalotes* tree. And while only about half of *Cephalotes* nodes were recovered in the analysis, we suspect that the high level of similarity among microbiota across the genus limits the overall level of support in the beta diversity clustering analysis, as variability within species overlaps substantially with variability between species (Fig. 1.3). Repeating this analysis with more sensitive techniques, such as the Low-Error Amplicon Sequencing approach recently developed by Faith *et al.* (2013), would help to determine whether the lack of support for the remaining internal nodes is due to biological variation, or is an artifact of the relatively high level of technical error we observed in our dataset (see Appendix I). Future techniques exploring patterns of codiversification in particular microbial lineages will provide further context to these community-level trends.

The apparent impact of recent bacterial evolution does not mean that niche-driven ecology or environmental filtering does not also play a role in structuring the *Cephalotes* microbiota. Rather, our analysis suggests that these factors are unlikely to be driving the bulk of the observed phylogenetic correlation. As noted above, some differentiation among clades is still apparent even at lower clustering widths (Fig. A1.6), indicating that there have probably been some more substantial shifts in the composition of the *Cephalotes* microbiota. These shifts may be apparent in the branches leading to *C. atratus* and to the *C. pallens* clade, for which community similarity patterns reflected host relatedness primarily at wider OTU clustering thresholds. Future studies, perhaps incorporating shotgun metagenomic data, will help to clarify the functional significance of these changes.

The primary strength of the sensitivity analysis approach we present here lies in its ability to effectively visualize correlation with microbiota composition across particular nodes of the host phylogeny. In this respect, it is especially complementary to more frequently applied techniques, like Mantel tests, which assess the overall strength of phylogenetic correlation in the dataset, but not its distribution. This is especially apparent in the ape dataset, where phylogenetic correlation at internal nodes was much more sensitive to parameter choice than at terminal nodes.

The usefulness of this approach extends beyond exploring phylogenetic correlation. The sensitivity analysis we've performed here is really a formalization of what is implicitly done whenever one compares the results of different beta diversity measures. Different diversity measures emphasize different properties of the underlying data, and performing such comparisons has been recommended as a general practice for understanding ecological patterns (Anderson *et al.* 2011). While we consider support for phylogenetic grouping of samples, in principle the same technique could be applied to any sample grouping hierarchy, allowing quick examination of support for a given hypothesis.

Social transmission and the stabilization of mutualism

Theory suggests that mutualisms should be vulnerable to cheating (Sachs *et al.* 2004). Despite this, evidence for breakdown of mutualistic lifestyles is comparatively rare (Sachs & Simms 2006; Sachs *et al.* 2011). Several mechanisms have been proposed to explain the evolutionary stability of mutualism (Sachs *et al.* 2004; Archetti *et al.* 2011), of which two -- partner choice and partner fidelity -- are particularly relevant to coevolution in gut microbiota. Partner choice mechanisms limit cheating by detecting and favoring interactions with cooperators; partner fidelity links fitness outcomes of partners, favoring shared investment over cheating.

Gut microbes present a challenging problem for mutualism: microbial mutualists localized to the gut lumen must be maintained in the face of a constant influx of food-associated microbes, potentially limiting partner fidelity, and are physically distant from epithelial-associated immune factors often associated with partner choice mechanisms (Nyholm & Graf 2012). In insects, partner fidelity via vertical transmission is typically associated with obligate intracellular symbionts (Moran *et al.* 2008). Many of the invertebrates that are known to rely on extracellular microbes for defined benefits -- such as light production in squid (Nyholm & McFall-Ngai 2004) and nutrition in stinkbugs (Kikuchi *et al.* 2007; Matsuura *et al.* 2012) -- have highly specific associations with just one or a few microbial lineages, and enhance the efficacy of partner choice mechanisms by physically sequestering these microbes for part or all of their lifecycle. While vertebrates do maintain a very complex lumenal gut community, it has been suggested that the vertebrate adaptive immune system may have evolved in part as a partner choice mechanism for dealing with this complexity (McFall-Ngai 2007). Such an immune-mediated mechanism could be responsible for the broad support for the grouping of conspecific apes in our sensitivity analysis (Fig. 1.5b) by imposing a host-specific selective filter on the acquisition of gut microbes from the environment.

In insects, social transmission could function to maintain relatively complex gut mutualisms by augmenting partner fidelity, effectively playing a similar role for an entire microbiota to that of ovarial transmission for individual insect endosymbionts. The high similarities we observed among *Cephalotes* nestmate microbiota indicate efficient homogenization of the colony's gut microbiota, and young *Cephalotes* queens presumably inherit the gut microbiota of their mother colony with similarly high fidelity, passing these on to their own offspring in turn. While experiments are ongoing to characterize the mutualistic, commensalistic, or parasitic nature of the various players in the *Cephalotes* microbiota, our data show that partner fidelity may be sufficiently strong to result in phylogenetic correlation of a substantial fraction of the microbiota with the host across tens of millions of years -- and across one of the most significant Neotropical ant adaptive radiations.

Notably, the non-cephalotine ants in our study hosted microbiota that were quite divergent among nestmates (Fig. 1.3), suggesting that eusociality alone is not sufficient to generate this degree of homogeneity. Others have proposed anal trophallaxis (i.e. adult feeding from anal secretions) as a mechanism for transmission of microbes in *Cephalotes* (Wheeler 1984; Russell *et al.* 2009; Anderson *et al.* 2012). This behavior has been observed in *Cephalotes* and *Procryptocerus* (Wilson 1976; Wheeler 1984) and is reportedly rare among ants generally (Hölldobler & Wilson 1990), consistent with the much greater variance observed in the three unrelated outgroup genera in our study. Anal trophallaxis has also been shown to be critical for transmission of beneficial microbes in bumblebee colonies (Koch & Schmid-Hempel 2011) and termites (Kitade 2004; Köhler *et al.* 2012), both of which host microbiota that appear to be more stable across their respective phylogenies than are those of their more solitary relatives (Martinson *et al.* 2011, Colman *et al.* 2012).

Conclusions

Our results, the first to explore *Cephalotes* gut microbiota in the broader context of host evolution, demonstrate remarkable lineage-wide stability. Many of the members of this community appear to have been present since the diversification of the host genus in the Eocene, and perhaps since before it split with *Procryptocerus* in the Cretaceous (Price *et al.*, in review). Using a novel application of sensitivity analysis, we have shown that correlation between these microbiota and their hosts' phylogeny appears to be driven largely by relatively recent bacterial evolution, suggesting it may be the result of codiversification. Notably, our reanalysis of great ape microbiota showed a substantially different pattern: while we still see patterns consistent with codiversification, much of the phylogenetic correlation might be better explained by phylogenetically-correlated selective forces such as diet or immunity.

We have presented an approach here that enables us to look at whole-community dynamics, while permitting some insight into the potential underlying drivers. Future techniques capable of identifying patterns of codiversification between hosts and individual members of complex communities will help us to better understand the composition of these broad patterns of similarity, and perhaps provide additional insight into the processes of transmission and coevolution underlying the patterns.

Acknowledgements

We thank Antonio Coral, Lina Arcila Hernandez, Gabriel Miller, Kleber del Claro, and Flavio Camarota for assistance with field work; Shauna Price for providing early access to the *Cephalotes* phylogeny; and Andrew Richardson, Yoel Stuart, and the class of OEB 210 for helpful comments on an early draft of this manuscript; and four anonymous reviewers for constructive criticism. Brazilian material was obtained under Material Transfer Agreement # 08/2009 between the Universidade Federal de Uberlandia and Harvard University, and Peruvian material under permits 394-2009-AG-DGFFS-DGEFFS and 79-2008-INRENA-IFFS-DCB to MEF from the Peruvian Ministry of Agriculture. This work was supported by Putnam Expedition Grants from the Museum of Comparative Zoology to JGS, a grant from the Milton Fund and the Harvard Society of Fellows to DJCK, a Doctoral Dissertation Improvement grant from the National Science

Foundation DDIG 1110515 to NEP and JGS, NSF Graduate Student Research Fellowship DGE 1144152 to JGS, and NSF awards DEB 0842144 and IOS 0841756 to SP.

Data accessibility

Raw reads from this study are available on the NCBI Short Read Archive under the Project Accession Number PRJNA227767. Additionally, computer scripts and example data are deposited in the Dryad data repository under DOI doi:10.5061/dryad.023s6, and processed sequence reads and metadata are available in MG-RAST under project number 6249.

CHAPTER 2

The whale gut microbiome digests animal prey using pathways common in terrestrial herbivores

Coauthors: Annabel C Beichman, Joe Roman, Jarrod J Scott, David Emerson, James J McCarthy, Peter R Girguis

Abstract

Mammals host dense gut microbiomes of immense physiological consequence, but the determinants of diversity in these communities remain poorly understood. While diet appears to be the dominant factor, host phylogeny also seems to play an important, if unpredictable, role. Here, we show that baleen whales, despite a carnivorous diet, harbor microbiomes with surprising functional and taxonomic similarities to those of terrestrial herbivores. These similarities likely reflect a convergent role for fermentative metabolisms in processing insoluble polysaccharides, like chitin, that are abundant in the whales' diet. In other respects, including protein catabolism and essential amino acid synthesis, baleen whale microbiomes more closely resembled those of terrestrial carnivores. Our results demonstrate that these aspects of the microbiome can vary independently even given an entirely animal-derived diet, helping to clarify the mechanisms by which diet and phylogeny combine to shape microbial diversity in the mammalian gut.

Introduction

Mammals host gut microbiomes that are immensely important to the health, and likely fitness, of the host (Ley *et al.* 2008b; McFall-Ngai *et al.* 2013). While the composition of these microbial communities is largely determined by host diet (Ley *et al.* 2008a; Muegge *et al.* 2011), a substantial amount of variation is correlated with phylogeny: some mammals with diets atypical of their clade, such as the herbivorous panda bear, host communities that are taxonomically more similar to their close relatives than to other mammals with similar diets (Ley *et al.* 2008a; Zhu *et al.* 2011). The determinants of this ‘phylogenetic inertia’ are not yet well understood, in part due to a paucity of data from mammals, like pandas, whose diets differ from those of their evolutionary ancestors.

Notably, the cetaceans (whales and dolphins) eat only animals, but evolved from herbivorous terrestrial artiodactyls related to cows and hippopotamuses (Gatesy *et al.* 2013). Baleen whales, which include the largest animals known to ever exist, are filter feeders, largely consuming small schooling fish and crustaceans with chitinous exoskeletons. On land, the closest extant relatives of whales are herbivores, with diets rich in cellulose, a structural analogue of chitin. Whales also have multichambered foreguts similar to those of their ruminant relatives (Langer 2001; Gatesy *et al.* 2013). Together, these factors make the whale microbiome an important point of reference for understanding the roles of diet and phylogeny in structuring mammalian gut communities. Yet, to date, the diversity and functional potential of these communities remains unknown.

To characterize the community composition of cetacean gut microbiomes, we used high-throughput sequencing to examine fragments of the 16S ribosomal RNA gene from fecal samples of several baleen whales, toothed whales, and terrestrial mammals, and compared these to data from previously sequenced mammals (for detailed collections information, see Table A2.1).

Methods

Sample collection

Baleen whale fecal samples for this project were collected primarily in August of 2011 from whales feeding in the Bay of Fundy, located between New Brunswick and Nova Scotia, Canada. Fisheries and Oceans Canada granted permits to JR and JM to collect whale feces in the Bay of Fundy (license #325842). Collection methods have been described previously (Roman & McCarthy 2010); briefly, whales were located and identified visually, and nearby fecal samples collected as quickly as possible post-defecation using either a 150 μm mesh dip net or a cod-end plankton net. Approximately 10 mL of these fecal samples were transferred to sterile plastic tubes and placed immediately on ice. These samples were then frozen at -20° C upon return to shore (see Table A2.1 for full collections information).

These samples were complemented by a variety of fecal specimens from wild and captive marine and aquatic mammals. Feces from Pacific Humpback Whales (*Megaptera novaeangliae*) from Sitka Sound and the Seymour Canal in Southeast Alaska were provided by Prof. Jan Straley, collected under NOAA permits 474-1700-02 and 14122. One specimen, from the Atlantic white-sided dolphin (*Lagenorhynchus acutus*), was collected from the colon during necropsy of a beached wild individual (under FOC license #325842). Two samples each were collected from captive bottlenose dolphins (*Tursiops truncatus*) and Beluga whales (*Delphinapterus leucas*), from the Long Marine Lab at the University of California at Santa Cruz and Mystic Aquarium in Mystic, Connecticut, respectively. Captive toothed whale samples were collected immediately following defecation by target individuals and frozen at -20° C. Four fecal samples from wild Hippopotamus (*Hippopotamus amphibius*) were provided by Dr. Amanda Subalusky. These samples were collected from fresh dung pats and preserved in the field using RNALater at room temperature, transferred to 4° C upon return to the US, and kept at 4° C until DNA extraction.

Finally, fecal samples from wild and captive terrestrial mammals were collected to ground truth our novel specimen types against previously published data for terrestrial mammals, as well as to serve as internal controls. Fresh scat from wild carnivores and herbivores was collected by members of the volunteer wildlife tracking group Keeping

Track MA. Additional specimens from captive herbivores were collected by the authors. All of these terrestrial samples were frozen at -20° C as soon as possible after collection, and remained frozen until DNA extraction.

DNA extraction and sequencing

DNA was extracted from samples using MoBio PowerSoil extraction kits under manufacturer's protocols. Approximately 40-200 mg of fecal material was used from each sample. Extracted DNA was quantified in a fluorometric assay using a Qubit fluorometer.

16S rRNA community profiles were characterized using Illumina MiSeq sequencing of the V4 region, and for a subset of samples, using 454 pyrosequencing of the V1–V3 regions. (Sequencing both regions permitted comparison to a broader range of publicly available datasets.) PCR amplification and Illumina sequencing of the 16S V4 region were performed at Argonne National Laboratories (Lemont, IL), following previously published protocols (Caporaso *et al.* 2011; 2012). Briefly, 1 µL aliquots of extracted DNA were amplified in triplicate 25 µL PCR reactions, using barcoded universal fusion primers 515f and 806r, at a 60° C annealing temperature, for 35 cycles. Triplicate aliquots were then combined per sample, normalized to equimolar concentrations, and pooled for sequencing on an Illumina MiSeq sequencer using paired-end, 250 base pair reads. PCR amplification and 454 pyrosequencing of the V1–V3 regions were both performed at Research and Testing Laboratories (Lubbock, TX) using previously published protocols (Dowd *et al.* 2008). Briefly, singlicate amplifications were performed in 25 µL reactions, using universal fusion primers 27F and 515R (Kumar *et al.* 2011), an annealing temperature of 54° C, and 1 µL of template. Normalized equimolar concentrations of PCR product were then pooled and sequenced using 454 Titanium chemistry (Roche).

For a subset of samples, we also performed shotgun metagenomic sequencing. We initially prepared libraries for six cetacean specimens (JS1, JS2, F5, F8, F12, and F16) using the PrepX ILM automated DNA library prep kit (WaferGen Biosystems; Fremont, CA) and NextFLEX Illumina-compatible barcode adapters (BIOO Scientific; Austin, TX) according to the manufacturer's recommended protocol, with starting input DNA quantities ranging from 2.5 to 18 ng, and size selection post-ligation via automated double-ended magnetic bead cleanup. Prior to library preparation, samples were sonically sheared in 50 µL of water using a Covaris S220 focused-ultrasonicator tuned to 400bp fragment size (Covaris; Woburn, MA). Size-selected libraries were then amplified for 14 cycles using NEBNext High-fidelity polymerase according to manufacturer's protocols (NEB; Ipswich, MA), except that amplifications took place in duplicate 25 µL reactions. Post-amplification, libraries were purified by hand using magnetic beads (Agencourt AMPure). Amplified libraries were then quantified via Bioanalyzer (High Sensitivity DNA assay, Agilent Technologies; Santa Clara, CA) and qPCR (KAPA Library Quantification Kit; KAPA Biosystems; Wilimington, MA), pooled in equimolar concentrations, and sequenced on an Illumina HiSeq 2500 instrument using paired-end 150 base pair chemistry.

To extend our internally-generated metagenomic dataset to a wider array of taxa, and to better match the longer read length available for previously published metagenomic datasets, we also sequenced an additional 14 libraries using Illumina MiSeq paired-end 250 base pair chemistry. These libraries included seven of our terrestrial or aquatic samples (JS3, JS4, JS5, JS6, JS7, JS8, and JS19), five additional cetacean samples (JS10, JS13, JS17, F9, and F11), and two technical replicates of samples previously sequenced (F12 and F16). Prior to library preparation, samples were sonically sheared in 50 µL of water using a Covaris S220 focused-ultrasonicator tuned to 400bp fragment size.

All 14 of these libraries were prepared using the KAPA LTP library preparation kit for Illumina and NextFLEX Illumina-compatible barcode adapters (BIOO Scientific) according to the manufacturer's recommended protocol, with starting input DNA quantities ranging from 8.4 to 180.6 ng, and size selection post-ligation via double-ended magnetic bead cleanup. Size-selected libraries were amplified for either 8 (JS3, JS4, JS10, JS19) or 10 (remainder) cycles using KAPA HiFi DNA polymerase amplification kits according to manufacturer's protocols, except that amplifications took place in duplicate 25 μL reactions. Amplified libraries purified by hand using magnetic beads (Agencourt AMPure), quantified via Bioanalyzer (High Sensitivity DNA assay, Agilent Technologies) and qPCR (KAPA Library Quantification Kit; KAPA Biosystems), pooled in equimolar concentrations, and sequenced on an Illumina MiSeq instrument using paired-end 250 base pair chemistry.

16S community sequence analysis

To maximize our ability to place the whale microbiota in the broader context of mammalian gut community diversity, we focused our analysis on Illumina sequences from the V4 region of 16S, for which the greatest diversity of mammalian microbiota samples were available at the time of analysis. From the Earth Microbiome Project database, we downloaded Illumina V4 datasets for a variety of myrmecophagous mammals (Delsuc *et al.* 2014) and for the subset of the mammals from Muegge *et al* (2011) that were resequenced at this region for the study by Delsuc *et al* (2014). Because a number of the myrmecophagous mammalian samples from Delsuc *et al* (2014) were identified as being potentially contaminated by environmental bacteria, these samples were removed prior to further analysis. Since the sequences from the Earth Microbiome Project were limited to the first 100 bases of the forward read, we trimmed our own sequences to match. After demultiplexing in QIIME, we concatenated sequences for all three datasets and proceeded with analyses.

Sequences were *de novo* clustered at 97% sequence identity and chimeras removed using UPARSE, following the procedure recommended by Edgar (Edgar 2013). Chimeras were identified using both *de novo* detection and using the Greengenes database (12_10 release) as a reference (Desantis *et al.* 2006). Following OTU picking, a QIIME-compatible OTU map was created using a custom python script; all subsequent analytical steps were performed in QIIME version 1.8.0 (Caporaso *et al.* 2010b). Clustered, putatively nonchimeric sequences were assigned taxonomies using the RDP classifier (Wang *et al.* 2007) trained on the aforementioned Greengenes database at an 80% confidence level. Representative sequences from each OTU were aligned to Greengenes and filtered of hypervariable positions using PyNast (Caporaso *et al.* 2010a), and a *de novo* phylogenetic tree was computed using FastTree (Price *et al.* 2009) within QIIME. For downstream analyses, OTU tables were rarified to 10,000 sequences, retaining the four samples (JS2, JS3, JS4, JS12) with lower sequence counts. Unweighted UniFrac pairwise distance matrices were calculated in QIIME, and visualized using NMDS ordination with the vegan package in R 2.15 (Oksanen *et al.* 2012; R Development Core Team). To enable comparison with a previously-published dataset of wild seal microbiota (Nelson *et al.* 2012), we downloaded that dataset from MG-RAST and repeated the above steps for 454 sequences from the V1–V3 regions of 16S.

Some studies have found the retention of DNA from food-derived bacteria in fecal samples (David *et al.* 2014). To help ensure that the sequences derived from whale feces were not in large part derived from their food organisms, we screened all samples analyzed in this study against a library of 231 full-length bacterial 16S sequences derived from calanoid copepods in the North Atlantic (Gerdts *et al.* 2013), the major food source of the right whales in our study. The representative sequence from each 97% OTU was searched against the copepod database using BLAST with an e-value cutoff of 10^{-30} and a percent identity cutoff of 97%. Hits were inspected to ensure the returned alignments were at least 75 base pairs in length.

We used both OTU and taxonomy-based approaches to identify bacterial lineages that were enriched in whales relative to other mammals. For OTU-based approaches, we utilized Kruskal-Wallace and nonparametric pairwise T significance tests in QIIME to identify specific *de novo* OTUs that were distributed significantly differently among categories. For taxonomy-based approaches, we summarized the rarified OTU table by family-level taxonomy in QIIME, then imported relative abundances from this table into the Galaxy implementation of LEfSe (Segata *et al.* 2011).

Shotgun metagenomic sequence analysis

We used analysis of shotgun metagenomic sequences to explore differences in functional capacity between gut microbiomes of whales and other mammals. To maximize the breadth of possible comparisons, we combined our data with previously published mammalian microbiome data from Muegge et al (2011), available on the MG-RAST database (Meyer *et al.* 2008).

To minimize the potential effect of differences in sequence processing between datasets, we ran our metagenomic sequences through the MG-RAST pipeline before analysis. After removing adapter contamination with CUTADAPT (Martin 2011), demultiplexed sequences were uploaded to MG-RAST using recommended settings for paired-end sequence data. After processing, filtered, dereplicated predicted open reading frames (files labeled *.299.* in the MG-RAST naming hierarchy), and translated amino acid sequences (labeled *.350.*) were downloaded for both of our datasets (HiSeq and long-read MiSeq) as well as the Muegge 454 dataset.

We used the HMP Unified Metabolic Analysis Network program (Abubucker *et al.* 2012) to estimate normalized abundances of metabolic genes and pathways. For all samples, predicted amino acid sequences were compared against the KEGG database release 62 (Kanehisa & Goto 2000) using the Ublast implementation of Usearch version 7.0.1001 (Edgar 2010), utilizing an acceleration parameter of 0.5 and an e value cutoff of 10^{-5}. For the six more deeply sequenced HiSeq samples, a random subset of 105,191 sequences were tested (equal to the number of predicted coding sequences in the most deeply sequenced sample from the Muegge dataset). Some samples showed signs of heavy contamination from host genomic sequences, notably the toothed whales (Fig. 2.S16). As a lack of available cetacean reference genomes precluded read mapping as a method for removing this contamination, we simply excluded reads for which the top Ublast hit was derived from a eukaryote.

Profiles of normalized, non-eukaryotic KEGG gene abundances were analyzed using the Galaxy implementation of LEfSe and in R version 2.15 (Segata *et al.* 2011; R Development Core Team). KEGG genes and pathways significantly enriched in whales relative to other mammals were calculated in LEfSe. To visualize similarity among sample categories, we used redundancy analysis in the vegan statistical package in R (Oksanen *et al.* 2012) to construct ordinations of KEGG genes and gene categories. Ordinations were calculated for the dataset as a whole, as well as for subsets of genes and pathways defined by higher levels of organization in the KEGG gene ontology. For two sets of metabolic genes that have previously been identified as significant differentiators of herbivorous and carnivorous diets—namely, central pyruvate metabolism and central glutamate metabolism (Muegge *et al.* 2011; David *et al.* 2014)—we manually curated lists of KEGG gene identifiers. We then used ANOVA and Tukey's post-hoc tests to determine significant differences between host categories.

Analysis of carbohydrate-active enzymes

Predicted proteins were annotated following previously published methods (Suen *et al.* 2010). First, unassembled metagenomic datasets were annotated against the carbohydrate active enzyme (CAZy) database (Cantarel *et al.* 2009) as follows. A local database of all predicted proteins corresponding to each family from the CAZy online database (http://www.cazy.org/, accessed: 01/27/2014) was constructed and used to align predicted proteins using BLASTP (v 2.2.28+) (e-value cutoff of 1e-05). Predicted proteins were then annotated against the protein family (Pfam) database (Punta *et al.* 2012) (ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd/, accessed: 01/27/2014) using RPSBLAST (v 2.2.28+) (Marchler-Bauer *et al.* 2011) (e-value cutoff of 1e-05). A CAZy-to-Pfam correlation list was compiled based on the secondary annotations provided through the CAZy online database. Only those proteins that had significant BLAST hits to a protein from our local CAZy database and its corresponding Pfam were retained and designated as a carbohydrate-associated enzyme. If no corresponding Pfam was identified, only CAZy hits with a bit score greater than 60 were retained. Next, we generated a table with rows corresponding to samples and columns corresponding to each CAZy family. Because some proteins may encode for several CAZy families (Cantarel *et al.* 2009), we allowed multiple CAZy annotations for individual proteins where appropriate. This table was quantile normalized using the default parameters in the metagenomeStats package in R version 3.0.2. For heatmap visualization and clustering, we applied Pearson's correlation coefficient index to the normalized data, clustered with average linkage UPGMA clustering, and produced the figure using the gplots package in R. Significance tests were performed on the normalized data matrix using the Galaxy implementation of LEfSe, with a p-value cutoff of 0.05 and a minimum effect size of 3.

Results and Discussion

Like most terrestrial mammals, baleen whales were dominated in large part by bacteria in the phyla Bacteroidetes and Firmicutes (Fig. 2.1a), though with a consistently higher proportion of reads assigned to the phyla Firmicutes and Spirochaetes and the classes Bacteroidia and Clostridia (see Table A2.1). Baleen whales hosted either no or very few reads assigned to Proteobacteria, the Bacilli, the genus *Coprobacillacus* within Erysipelotrichaceae, or the genera *Blautia* and *Coprococcus* within the Lachnospiraceae, all of which were comparatively common among terrestrial mammals. Some of these lineages were especially depleted in the right whales compared to the other baleen whales (humpback and sei whales), which also grouped closer to terrestrial mammals on a PCoA ordination of UniFrac distances (Fig. 2.1b). Suggestively, these two groups of whales consume substantially different diets, with right whales especially dependent on comparatively chitin-rich calanoid copepods; these differences are discussed further below.

Several lineages that were comparatively abundant in baleen whale samples were also relatively enriched in terrestrial herbivores, including the clade 5 Verrucomicrobia, the phylum Lentisphaerae, the clade RF3 Tenericutes, and the genus *Treponema* in the Spirochaetes. Notably, these groups were all abundant in both ungulate and non-ungulate terrestrial herbivores, suggesting their shared presence in whales may be due as much to functional convergence as to phylogenetic conservatism. The toothed whale microbiota in our sample were highly variable in composition and are discussed further elsewhere (see Appendix 2).

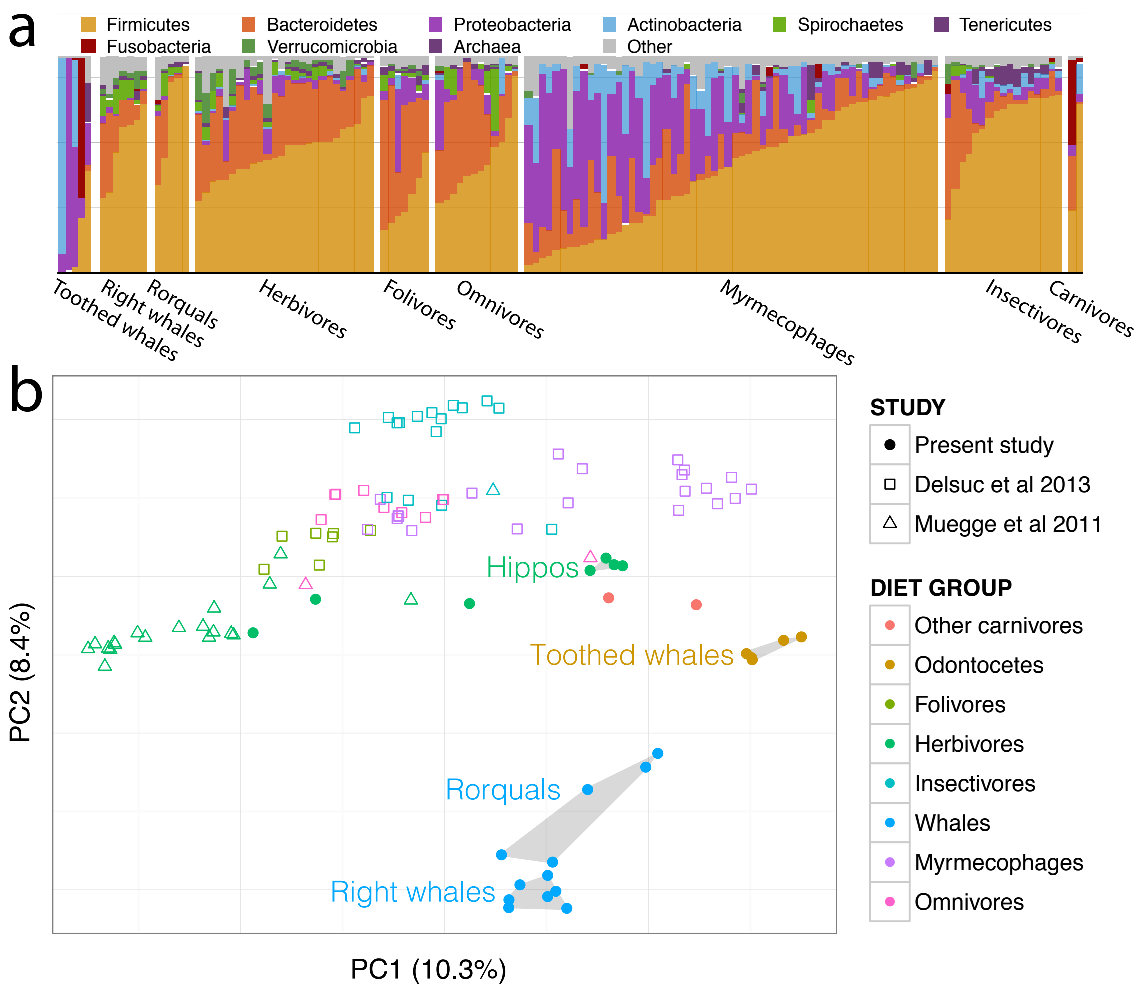


Figure 2.1: Baleen whales host a distinct microbiota. (a) Taxonomic composition of 16S rRNA amplicon sequences in cetaceans and terrestrial mammals. See Fig S2 for additional detail. (b) PCoA ordination of unweighted UniFrac distances among mammalian gut microbiota.

Although the broad taxonomic composition of the baleen whale microbiota was similar to those of terrestrial mammals, it was strikingly different at the level of 97% similarity Operational Taxonomic Units (97% OTUs, Fig. 2.1b). Of 397 OTUs with an average abundance of at least ten read per whale, 87.9% were significantly different in abundance when compared to terrestrial mammals. As with terrestrial carnivores, our whale samples had comparatively low diversity (Fig. A2.2). Targeted exploration of the most abundant baleen whale OTUs suggests that the differentiation from observed terrestrial OTUs may have derived from a long separation from the terrestrial environment rather than acquisition from the marine environment: the top BLAST hits to these sequences are primarily to terrestrial animal gut bacteria. We also found little evidence of sequences from food or the environment in baleen whale samples: OTUs that had a 97% identity to a library of bacterial 16S sequences from North Atlantic copepods, the preferred food of right whales, accounted for six or fewer of the 10,000 rarified sequences in each of the baleen whale samples (Fig. A2.3).

Given the whales' animal-based diet, the taxonomic affinities between baleen whale and terrestrial herbivore microbiota were particularly unexpected. To help establish whether these phylogenetic similarities have functional underpinnings, we sequenced and analyzed the predicted coding content of shotgun metagenomes for eight baleen whales. Our results showed that the whale microbiome echoes the carnivore microbiome for genes related to protein digestion and biosynthesis, but has surprising similarity to herbivore microbiomes with respect to genes involved in carbon and energy metabolism —likely reflecting the shared relevance relevance of fermentation of polysaccharide compounds in their diet.

In general, the functional potential of a microbiome has been reasonably represented by the patterns observed in 16S-based community profiles (Muegge *et al.* 2011) (though 16S diversity may more poorly predict specific subsets of functional potential (Cantarel *et al.* 2012)). Consistent with this, overall functional composition in our dataset reflected the patterns observed in 16S-based community profiles (Fig. 2.2a). In ordinations of predicted KEGG pathway abundances (Fig. 2.2a) and gene abundances (Fig. A2.4a), the baleen whales formed a cluster distinct from terrestrial mammals, with the humpback and sei whales in our metagenomic sample set slightly closer to terrestrial mammals, on average, than were the right whales.

Although baleen whales grouped independently from terrestrial mammals when considering the entire set of KEGG genes or pathways, they tended to cluster with different terrestrial dietary groups at lower levels of the KEGG hierarchy. For example, baleen whale microbiomes were relatively similar to those from terrestrial carnivores in the abundance of pathways involved in essential amino acid metabolism and biosynthesis (multivariate effect size $\text{eta}^2 = 0.15$ vs carnivores and 0.71 vs herbivores, Fig. 2.2d), likely reflecting high nitrogen content in both diets. Both the baleen whale and terrestrial carnivore microbiomes were relatively depleted in enzymes catalyzing the final steps in biosynthesis of most essential amino acids (Fig. A2.5). Baleen whale microbiomes were also enriched in genes catalyzing the degradation of glutamine and glutamate, and depleted in genes catalyzing the synthetic reactions (Fig. 2.2g). Glutamate metabolism is one of the most diagnostic pathways in distinguishing between microbiomes exposed to predominantly animal- and plant-based diets, both in humans (David *et al.* 2014) and in mammals generally (Muegge *et al.* 2011). Together, these genes suggest a predominantly catabolic direction of protein metabolism, reflecting the whales' animal-based diet.

In contrast to this largely "carnivorous" profile for amino acid-related genes, baleen whales were considerably more similar to terrestrial herbivores when considering pathways in the KEGG energy metabolism ($\text{eta}^2_{\text{carnivore}} = 0.69$, $\text{eta}^2_{\text{herbivore}} = 0.24$, Fig. 2.2b) and lipid metabolism ($\text{eta}^2_{\text{carnivore}} = 0.56$, $\text{eta}^2_{\text{herbivore}} = 0.31$, Fig. 2.2c) categories. Like glutamate metabolism, pyruvate metabolism has been shown to be a key difference separating herbivorous and carnivorous microbiomes (Muegge *et al.* 2011; David *et al.* 2014). Similar to terrestrial herbivores, baleen whales were enriched in genes withdrawing metabolic intermediates from the TCA cycle, and depleted in genes catalyzing the reverse reactions (Fig. 2.2f).

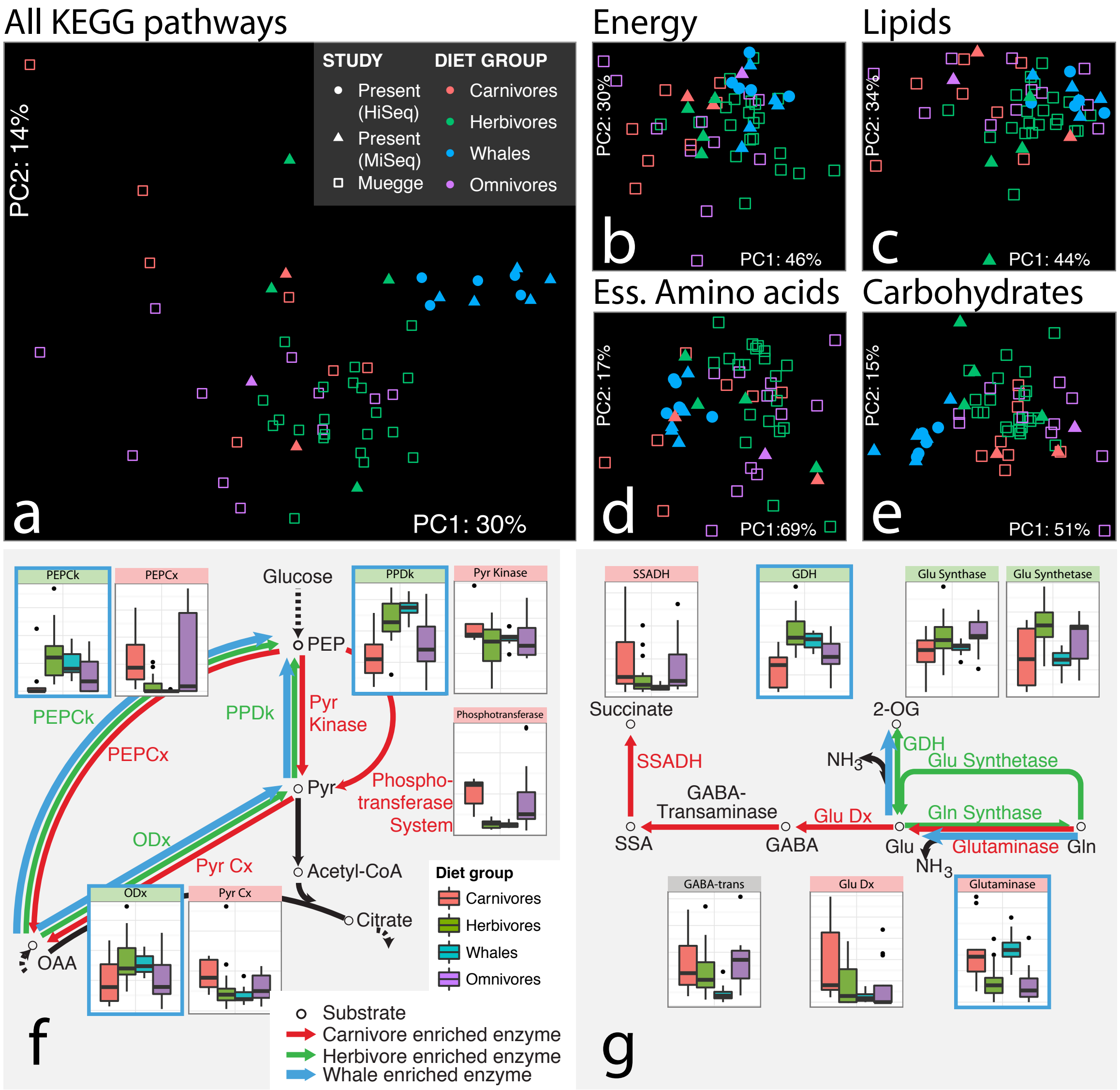


Figure 2.2: The functional compositions of baleen whale microbiomes show similarity to those of both terrestrial herbivores and carnivores. PCA ordinations of predicted metagenomic functional potential show baleen whales microbiomes are distinct from those of terrestrial mammals when considering all pathways (a) or pathways involved in carbohydrate metabolism (e). Pathways involved in energy (b) and lipid (c) metabolism are more similar to those of terrestrial herbivores, and pathways involved in synthesis of essential amino acids (d) are more similar to those of terrestrial carnivores. Sub-pathways previously shown to separate herbivore and carnivore metagenomes show a similar split, with baleen whale metagenomes showing a pattern of enrichment similar to herbivores in central pyruvate metabolism (f) and to carnivores in glutamate metabolism (g). Distributions of normalized abundances are shown as box plots for each gene, colored according to dietary category. Genes relatively enriched in terrestrial herbivores and carnivores have headers colored green and red, respectively. Those enriched in whale microbiomes are outlined in blue, and the proposed direction of metabolite flux for each dietary category given as colored arrows between metabolites. Pathway figure after Ref 4.

The surprising similarities in carbon metabolism between microbiomes of plant-eating terrestrial herbivores and animal-eating baleen whales may be explained by a shared reliance on fermentation. Like their herbivorous artiodactylate ancestors, whales possess a multichambered stomach (Langer 2001; Gatesy *et al.* 2013). Baleen whales also possess a blind-end caecum between the ileum and colon, which is not present in the toothed whales (Langer 2001). Given their morphological similarities with herbivores, questions about the mechanisms and importance of chitin digestion in baleen whales have been outstanding for decades (Herwig *et al.* 1984; Olsen & Mathiesen 1996). One study found short chain fatty acids (SCFAs), the presumed end products of microbial fermentation, in grey and bowhead whale forestomachs that were comparable in concentration and composition to those in the forestomachs of terrestrial ruminants (Herwig *et al.* 1984). Yet the importance of fermentation for whales has been contested, with others (Olsen & Mathiesen 1996) arguing that the rates of SCFA production they measured would only account for a small fraction of the animal's daily energy budget. Similarly, there is mixed evidence in whale guts for methane production, a process commonly associated with fermentation in terrestrial animals (Herwig *et al.* 1984; Hackstein & van Alen 1996; Olsen & Mathiesen 1996).

Although it is impossible to draw strong inferences about metabolite flux solely from potential metagenomic capacity, our dataset in conjunction with these previous measurements of SCFA production support a major role for fermentation in the whale gut microbiome. Baleen whale microbiomes were considerably more similar to those from terrestrial herbivores than to those from terrestrial carnivores in the abundance of genes related to the metabolism of pyruvate, a major fermentation intermediate. Although the causal explanation for this difference in the microbiome-wide directionality of pyruvate metabolism has not been well documented, the reduced role of oxidative phosphorylation in fermentative microbes may make replenishment of TCA intermediates less important. Consistent with measurements of high concentrations of SCFAs in baleen whale foreguts, we also found that both whale and herbivore microbiomes were enriched in enzymes catalyzing the production and utilization of SCFAs (Fig. A2.6). Excess hydrogen produced during fermentation in terrestrial herbivore guts is consumed in methanogenesis or acetogenesis via the Wood-Ljungdahl pathway, and most major enzymes in this pathway were enriched in both herbivores and baleen whales relative to carnivores (Fig. A2.7a). Finally, we found higher proportional abundance of the key methanogenesis enzymes heterodisulfide reductase and methyl-coenzyme M reductase in both baleen whales and herbivores, suggesting that these whales do indeed retain the capacity for methanogenesis (Fig. A2.7b). Our whale samples were devoid of 16S amplicons allied to the Methanobacteria, the group of Archaea classically associated with methanogenesis in mammal guts, and which we recovered in some abundance from our terrestrial herbivore samples. However, both the baleen whales and terrestrial herbivores had a substantial number of reads related to the recently described Euryarchaeal order Methanomassiliicoccales (Dridi *et al.* 2012; Paul *et al.* 2012). These organisms belong to a clade of obligate hydrogen-consuming methanogenic methylotrophs found so far exclusively in gut habitats, including humans, termites, and cattle, and have recently been shown to be important methane producers (Poulsen *et al.* 2013). Notably, some of these organisms can utilize methylated amines, which are abundant solutes in marine zooplankton (Strøm 1979), as a substrate (Poulsen *et al.* 2013).

Baleen whale microbiomes grouped independently from the terrestrial mammals only in the category of carbohydrate metabolism, likely reflecting fundamental differences in the source of carbohydrates (Fig. 2.2e). Fermentation in terrestrial herbivores is fed largely by plant-derived cellulose and hemicellulose, while the predominant carbohydrate source in many baleen whales is the chitinous exoskeleton of invertebrate prey. Differences in these feedstocks are be reflected in the monosaccharide kinase profiles of these groups. Terrestrial herbivore microbiomes are enriched in kinases for galactose, rhamnose, and xylulose, three major constituents of hemicellulose that are absent in chitin (Fig. A2.8). By contrast, whale, terrestrial herbivore, and terrestrial carnivore microbiomes were all similarly enriched in gluco- and fructokinases.

To further examine carbohydrate utilization in the baleen whale gut microbiome, we used a combination of sequence similarity and hidden Markov model searches to generate profiles of carbohydrate-active enzymes (CAZymes) for whale and terrestrial microbiomes. Baleen whales hosted a unique complement of these genes, consistent with the hypothesis that fermentation of animal carbohydrates is especially important in this community. UPGMA clustering of CAZy abundance profiles showed baleen whales grouping separately from terrestrial mammals, but nested within the carnivores rather than herbivores—reflecting the shared importance of animal polysaccharides in whale and terrestrial carnivore diets, despite the whales' similarities to herbivores in downstream fermentation pathways (Fig. 2.3). Among the 179 CAZy families represented in the total dataset, 36 were significantly more abundant in baleen whale metagenomes than in those of terrestrial mammals. These families are predominantly associated with the degradation of animal-derived polysaccharides. Nine of the significantly enriched families are described to have activity on chitin and associated compounds, including enzymes associated with chitin binding, hydrolysis of long-chain chitin polymers (chitinases), and hydrolysis of chitin oligomers (chitobiases). Chitin-related genes were also especially abundant as an overall fraction of carbohydrate-active genes. The chitobiase GH20 was on average the third most abundant CAZyme in the baleen whale metagenomes, and was the most abundant of the enriched CAZymes (Fig. 2.3). Nearly as abundant was the carbohydrate binding module CBM37, which has been shown to bind to a number of substrates including chitin (Xu *et al.* 2004), and which has been proposed to function in the localization of glycanases to the bacterial cell wall for attachment to and degradation of extracellular polysaccharides (Ezer *et al.* 2008). Together with the chitinase GH18, these CAZymes comprised three of the six most abundant families enriched in baleen whales.

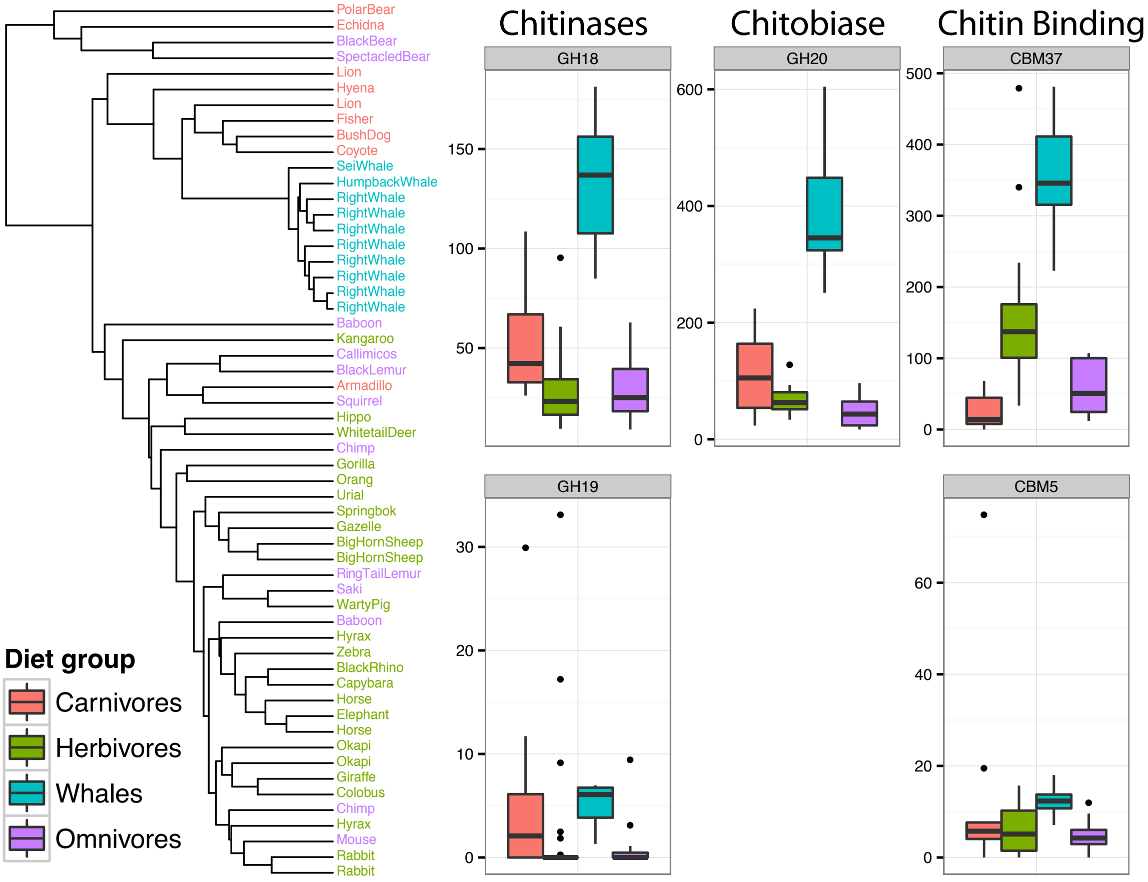


Figure 3: The composition of carbohydrate-active enzymes (CAZymes) in baleen whale microbiomes is distinct and enriched in genes predicted to have activity on chitin. (left) UPGMA clustering dendrogram of CAZyme abundances. Dietary compositions are indicated by tip label color. (right) Box plots showing distribution in normalized abundance of the five most abundant chitin-related CAZymes significantly enriched in baleen whale relative to terrestrial microbiomes. Note that CBM5 and CBM37 are binding domains rather than enzymes, and CBM37 has activity against a broad spectrum of polysaccharides in addition to chitin (Xu *et al.* 2004).

The distribution of CAZy families in the whale gut likely reflects one of the major axes of variation in both taxonomy and function relative to other mammalian microbiomes. Profiles of CAZy families have been found to be strong predictors of diet in mammalian and human microbiomes (Eilam *et al.* 2014), and different microbial lineages seem to be consistently associated with particular carbohydrate categories. Bacterial isolates in the genus *Bacteroides*, for example, have thus far been found to encode especially large numbers of animal-carbohydrate active enzymes, while isolates from the related genus *Prevotella* have encoded only plant-active enzymes (Kaoutari *et al.* 2013; Eilam *et al.* 2014). Consistent with this pattern, our baleen whale samples were relatively enriched in 16S sequences assigned to the genus *Bacteroides*, but completely devoid of sequences assigned to the genus *Prevotella* in the family Prevotellaceae. However, the profile of CAZyme diversity within bacterial lineages appears to be quite labile (Cantarel *et al.* 2012), and more targeted work will be necessary to identify the functional roles of specific bacteria within the whale gut.

The surprising affinities between the microbiomes of terrestrial herbivores and the baleen whales in this study help to clarify the complex interplay between the dietary and physiological determinants of the mammalian gut microbiome. As in previous surveys of mammalian gut microbiota, our data show correlations with both diet and host phylogeny (Ley *et al.* 2008a; Muegge *et al.* 2011), depending on which portions of the data are being considered. But in baleen whales, the correlation between host phylogeny and microbiome composition may reflect constraints imposed by the physical structure of the gastrointestinal tract itself, with the multichambered artiodactyl foregut serving as a preadapted fermentation chamber. Under the model proposed here, it is this morphological preadaptation that set the stage for the degradation of chitin – a polysaccharide with structural similarities to cellulose, but requiring an entirely different set of carbohydrate-active enzymes for digestion—to drive a fermentative profile for overall metabolism in the microbiome that is in many respects similar to that observed in the whales' terrestrial herbivore ancestors.

A multichambered foregut may also help to explain the lack of affinity we observed between the microbiota of baleen whales and those of terrestrial insectivores and myrmecophages. These carnivorous mammals also eat chitin-rich arthropod-based diets; but unlike whales, they have relatively simple guts. Future surveys of the functional potential of insectivorous mammalian microbiomes should help to further clarify the role of host gut physiology in modulating the microbiome's response to dietary input (Delsuc *et al.* 2014).

Such a preadaptation may also have played an important role in the evolution of the baleen whales by facilitating the maintenance of a microbial community capable of extracting recalcitrant nutrients from a zooplankton diet. Although empirical estimates of the microbial contribution to the total proportion of whales' caloric intake vary (Olsen & Mathiesen 1996), chitin alone may comprise as much as 3 - 7% of the total dry weight of krill and copepods (Vijverberg & Frank 1976; Nicol & Hosie 1993). Baleen is a relatively recent innovation, and ancestral whales were likely piscivorous (Gatesy *et al.* 2013). Consequently, the whales' most recent ancestor may have had gut microbiomes more similar to modern-day dolphins and seals (Nelson *et al.* 2013). More extensive sampling of toothed whales (see Appendix 2) will help to further constrain the relative roles of diet, environment, and phylogeny in structuring the diversity of mammalian gut microbiomes. Finally, our results highlight the potential impact of the multichambered artiodactyl gut on global marine biogeochemistry. Using many of the same pathways and microbes that their terrestrial relatives employ to digest cellulose, the most abundant biopolymer on land, whales evolved a process to digest chitin, the most abundant biopolymer in the sea (Beier & Bertilsson 2013). Given its role as a major marine reservoir of both carbon and nitrogen, broad questions about the distribution of chitin-degrading microbes in the sea have been pursued for over 75 years (Zobell & Rittenberg 1938). The digestive capacity of baleen whales has consequences for elemental flux throughout the ocean, including enhanced benthic-pelagic coupling; increased marine productivity as a result of the near-surface release of nutrient-rich fecal plumes; the transfer of nutrients to areas of low productivity during migration; and organic enrichment of the deep sea via whale carcasses (Roman *et al.* 2014). Such processes may owe as much to the microbes in the belly of the whale as to the whales themselves.

Acknowledgments

Field research was supported by a grant from the Marine Mammal Commission to JR and JJM. Samples were collected under Fisheries and Oceans Canada license # 325842. Additional support was provided by the Museum of Comparative Zoology and the National Science Foundation (IOS-1257755 and OCE-0732369 to PRG; OCE-1155754 to DE). We thank John Nevins for field and laboratory assistance, advice, and project coordination; Heather Koopman, Andrew Westgate, Laurie Murison, Jared Juckiewicz, and Zach Siders for enabling us to collect samples in the Bay of Fundy; and Robert Wayne, Naomi Pierce, Andrew Berry, Leonora Bittleston, Chris Baker, Jennifer Delaney, and Steven Worthington for discussion and support. Jan Straley, Amanda Subalusky, Terrie Williams, Aleksija Neimanis, Tracy Romano, the Mystic Aquarium, and Keeping Track MA supplied additional fecal samples.

Data Accessibility

Sequence data generated for this project have been deposited in the MG-RAST database under the project accession mgp3854.

CHAPTER 3

Gut bacterial densities help to explain the relationship between diet and habitat in rainforest ants

Co-authors: Piotr Lukasik, Megan E Frederickson, Naomi E Pierce

Abstract

Modern amplicon-based sequencing techniques permit the rapid assessment of diversity in microbial communities, and have consequently led to an explosion in data describing the distribution of microbial taxa across environments. However, these techniques by themselves typically provide no information about the absolute abundance of microbes. Abundance is a key parameter in ecology, and important to estimates of potential metabolite flux, impacts of dispersal, and sensitivity of samples to technical biases such as laboratory contamination. Here, we use fluorescence microscopy and quantitative PCR as independent estimates of microbial abundance to test the hypothesis that microbial symbionts have enabled ants to dominate tropical rainforest canopies by facilitating herbivorous diets. Through a systematic survey of ants from a lowland tropical forest, we show that the density of gut microbiota varies across several orders of magnitude among ant lineages, with median individuals from many genera only marginally above detection limits. Supporting the hypothesis that microbial symbiosis is important to dominance in the canopy, we find that the abundance of gut bacteria is positively correlated with stable isotope proxies of herbivory among canopy-dwelling ants, but not among ground-dwelling ants. Furthermore, we report evidence of midgut bacteriocytes in two unrelated ant genera from which they were previously unknown, suggesting that the incidence of this symbiotic habit among ants is much higher than has been thought. Our results help to resolve a longstanding question in tropical rainforest

ecology, and have broad implications for the interpretation of sequence-based surveys of microbial diversity.

Introduction

When tropical entomologists began systematic surveys of arthropod biomass in rainforest canopies, the dominance of ants in the fauna appeared to be paradoxical. As formulated by Tobin (1991), the problem centered around an apparent inversion of the classic terrestrial ecosystem biomass pyramid: ants were presumed to be predators or scavengers, yet frequently outweighed their putative prey. This biomass "paradox" (Davidson & Patrell-Kim 1996) was partly resolved by evidence from stable isotope analysis that most canopy ants are functionally herbivorous (Cook & Davidson 2006; Douglas 2006; Davidson:2003hh; see also Eilmus & Heil 2009). These ant herbivores feed to a large extent on plant-derived liquid foods, including extrafloral nectar and hemipteran exudates (Davidson *et al.* 2004). But the limited availability of nitrogen in these resources itself poses a dilemma: how do herbivorous canopy ants acquire nitrogen resources that are both abundant and balanced enough in amino acid profile to sustain colony growth?

Insects with nutrient-imbalanced diets almost ubiquitously rely on bacterial symbioses to complement their nutritional demands (Moran *et al.* 2008; Engel & Moran 2013). Indeed, evidence has been found for specialized associations between bacteria and a number of canopy ant lineages. *Blochmannia* bacteria were among the first described endosymbionts (Blochmann 1888), and appear to play a role in upgrading nitrogen for their host *Camponotus*, a frequent resident of forest canopies (Feldhaar *et al.* 2007). Billen and collaborators showed that several species of the old-world arboreal genus *Tetraponera* have a bacterial pouch at the junction between the mid- and hindgut which houses a dense community of extracellular bacteria (Billen & Buschinger 2000). Others have drawn connections between these bacteria and nitrogen-fixing nodulating bacteria of legumes, in part due to detection via PCR of the nitrogenase gene *nifH* (van Borm *et al.* 2002; Stoll *et al.* 2007); though to our knowledge, no nitrogenase activity has ever been demonstrated. Specialized extracellular bacteria have long been known to inhabit the morphologically elaborated guts of the new-world arboreal genus *Cephalotes* (Caetano & da Cruz-Landim 1985; Roche & Wheeler 1997; Bution *et al.* 2007). In *Cephalotes*, stability across and correlation with the host phylogeny suggest an important and conserved role for these microbes (Sanders *et al.* 2013), and experimental response to changes in diet suggest that role may relate to nutrition (Hu *et al.* 2013). Other ant lineages that have been surveyed, including the invasive fire ant, show less evidence for specialized bacterial associations (Lee *et al.* 2008; Meyer *et al.* 2011; Sanders *et al.* 2013).

Fewer studies have systematically surveyed bacteria across ants. One major comparative analysis, covering representatives of two-thirds of known ant genera, did detect a systematic relationship between herbivory (defined by stable isotope composition) and presence of an ant-specific lineage of alphaproteobacteria related to the genus *Bartonella* (Russell *et al.* 2009). As in *Tetraponera* (Stoll *et al.* 2007), Russell *et al.* were able to amplify the *nifH* gene from some of these ant specimens, though no nitrogenase activity was observed in acetylene reduction assays. But despite the lack of any direct evidence of a functional relationship, the distribution of ant-specific bacteria across lineages with independently-evolved herbivorous diets make a compelling case for a generalized role for bacteria in facilitating herbivory in ants.

Much of the research that has been done to describe insect-associated bacterial communities, especially since the advent of high-throughput next-gen sequencing, suffers from a common limitation: a lack of context as to the absolute abundance of the microbes being surveyed. PCR-based microbial community profiling techniques, including cloning and Sanger sequencing, restriction fragment polymorphism analysis, and next-gen amplicon sequencing, almost always start with an amplification step to produce many copies of the original template DNA. The resulting libraries retain almost no information about starting template abundance. The amplification step can be subject to contamination, especially for samples (like ants) with very low starting DNA concentrations (Salter *et al.* 2014). Even in the absence of contamination, the biological implications of very low density bacterial communities are likely to be substantially different than for symbionts, like the nutritional endosymbionts of *Camponotus* ants (Wolschin *et al.* 2004), that are present in very high numbers. Without additional information about absolute abundance, it is difficult to know whether novel microbial ecosystems are more akin to sparse deserts or dense rainforests.

Insects known to rely on bacterial symbionts for nutrient complementation also tend to support relatively high densities of those symbionts (Schmitt-Wagner *et al.* 2003; Martinson *et al.* 2012; Engel & Moran 2013). Here, we examined how bacterial abundance varies across ant species in a tropical rainforest, and whether more herbivorous ants support more bacteria. We used two independent methods—quantitative PCR and fluorescence microscopy—to assess absolute abundance of cells while also gaining insight into their localization and morphology. Our findings reveal surprising diversity in the nature and density of these associations, providing critical context for understanding the roles that microbes play in this major element of the rainforest ecosystem.

Methods

Field collections

We performed primary collections in July-August 2011 at the Centro de Investigaciones y Capacitacion de Rio Los Amigos (CICRA) in southeastern Peru. CICRA contains a mix of primary and secondary lowland tropical forest. We collected opportunistically from most available habitat types at the station, finding ants primarily visually but also using a mix of baits to recruit workers. To ensure that individuals came from the same colony, we took workers from within nests when possible; but when nests were inaccessible or could not be found, multiple workers were taken from the same foraging trails. In all cases, we brought live workers and/or nest fragments back to the field station for processing. Each colony was processed within 24 hours of collection.

When numbers allowed, we preserved tissues for nucleic acid analysis, stable isotope analysis, FISH microscopy, and morphology. First, workers were sacrificed by immersion in 97% ethanol. They were subsequently surface sterilized in 0.5% sodium hypochlorite solution for approximately 1 minute, then rinsed twice in sterile PBS buffer. For preservation of nucleic acids, the midgut and hindgut of worker ants were dissected with sterile forceps in clean PBS buffer and preserved in RNA Later, one ant per vial. One GI tract per colony was also completely dissected and visualized immediately using fluorescence microscopy (see below). The heads, legs, and mesosomas from these dissected individuals were preserved together in 95% ethanol in a separate tube for analysis of stable isotopes.

To preserve for subsequent fluorescence microscopy, we semi-dissected whole worker gasters to expose internal tissues. These were fixed in 4% PBS-buffered paraformaldehyde at room temperature for 2 hours or at 4° C over night, washed four times in sterile PBS solution, dehydrated using four washes with 75% ethanol:PBS solution, and preserved in molecular grade ethanol.

Any remaining workers were preserved whole in 95% ethanol for morphological identification.

In addition to the primary collection at CICRA, we collected a secondary set of specimens in August 2013 for additional FISH microscopy. These collections took place in secondary forest at the Villa Carmen field station in southern Peru.

Microscopy

For most colonies, we visualized a single dissected worker gut in the field using SYBR Green fluorescence microscopy. Guts dissected as above were placed on a glass slide, covered with a 1:100 mixture of SYBR Green and VectaShield mounting medium, and torn open using forceps to expose the contents of the midgut and ileum. Slides were covered with a glass coverslip and sealed with clear nail polish, then visualized on an AmScope epifluorescence microscope powered by a portable generator. Putative bacterial cells were identified by size and morphology, and the abundance estimated using a roughly logarithmic visual scale (0 = no visible bacterial cells, 1 = tens, 2 = hundreds, 3 = thousands, 4 = tens of thousands; see Fig. A3.1). Representative photomicrographs for each colony were taken with a digital camera.

Preserved ant tissues proved especially difficult to use for FISH microscopy relative to tissues from other insects, rapidly losing morphological structure when fixed in only acetone or ethanol, and displaying high levels of autofluorescence. For convenience, we have provided detailed protocols as supplemental material. Briefly, preserved tissues were bleached in a solution of hydrogen peroxide and ethanol for several days to decrease autofluoresence. Whole-mount specimens were rehydrated in buffer, dissected further if necessary, washed in hybridization solution, and hybridized with a solution containing FISH probes and DAPI. Hybridized samples were then washed and mounted in an antifade medium on a slide for visualization. Specimens for tissue sections were dehydrated in acetone before embedding in glycol methacrylate resin (Technovit), and 1-2 μm sections cut on a microtome. These sections were hybridized with a solution containing FISH probes and DAPI, then washed visualized under an antifade medium.

Nucleic acids analysis and quantitative PCR

We extracted DNA from individual dissected guts using the PowerSoil 96-well DNA extraction kit from MoBio. First, we added 1 volume of sterile molecular-grade water to tubes containing dissected guts to help redissolve any precipitated ammonium sulfate. Tubes were vortexed several times at room temperature until any visible precipitate had dissolved, then spun in a microcentrifuge at 10,000 xg for 10 minutes to pellet cells and tissues. We removed the supernatant and replaced it with 200μL buffer C1 from the PowerSoil extraction kit, vortexed at maximum speed for 15 seconds to resuspend tissues, and transferred this solution to the extraction plates. From there, we proceeded with the extraction according to the manufacturer's protocol.

We quantified extracted DNA using PicoGreen dsDNA quantification reagent (Thermo Scientific), following the manufacturer's protocol for 384-well microplate formats. Due to limited quantities of eluted DNA, the protocol was modified slightly: rather than mixing equal volumes sample and PicoGreen reagent solution, 10 μL of sample solution was added to 30 μL quantitation reagent, diluted correspondingly with molecular grade water. Each sample was measured in triplicate on the same 384-well microplate. Plates were read on a Spectramax Gemini XS fluorescence plate reader, and standard curves fit in SOFTmax PRO (Molecular Devices, Inc.). The mean of the three replicates was taken as the DNA concentration for each extraction.

PCR quantitation of bacterial 16S rRNA molecules was performed with SybrGreen chemistry (PerfeCTa SYBR Green SuperMix, Quanta Biosciences) using the primers 515F and 806R (Caporaso *et al.* 2011), each at 250 pM. This primer pair was chosen to permit direct comparison of qPCR values with Illumina-sequenced amplicons of the same locus.

Two microliters of extracted DNA were used per 20 μL amplification reaction in 96-well plates. Reactions were performed on a Stratagene MX3000p real-time thermocycler, using 40 iterations of the following three-step cycle: 45 seconds denaturation at 94° C, 60 seconds annealing at 50°, and 90 seconds extension at 72°. In addition, a 3 minute initial denaturation at 94° and a post-amplification denaturation curve were performed. To increase measurement accuracy, each sample was run at least twice, with each replicate occurring on a separate PCR plate. These technical replicates were averaged before further analysis. For absolute quantification, we included in triplicate a 1:10 serial dilution standard curve generated from linearized plasmids containing full-length E. coli 16S rRNA. Due to background amplification from 16S present in reagents, some amplification was observed at high cycle numbers in no-template controls (mean NTC amplification estimated at 85 copies / μL). The mean background amplification from three no-template controls per plate was subtracted from each sample on that plate, and samples below this limit of detection normalized to 1 copy per μL. To test for specificity of qPCR primers, we also prepared a standard curve of 18S rRNA amplified from an ant (*Cephalotes varians*).

Isotopic analysis

To estimate the relative trophic position of the ant colonies in this study, we analyzed ethanol-preserved tissues using stable isotope ratio mass spectroscopy. Heads and mesosomas from the individuals used for gut dissections were preserved in a separate vial of 95% ethanol to minimize the isotopic contribution of materials from the gut. For each colony analyzed, these tissues were dried overnight at 60° C, ground into powder with a mortar and pestle, and ~ 5 mg of powder placed in a silver foil capsule. These were combusted and analyzed for ∂15N at the Boston University Stable Isotope Laboratory.

Statistical analysis

We tested the correlation of both visual and quantitative PCR estimates of bacterial abundance using linear and generalized linear mixed models with the lme4 package in R. Because visual estimates corresponded roughly to a step function with respect to qPCR estimates (Fig. A3.3), we treated these data as presence / absence, with visual estimates of 0 or 1 corresponding to 'absent' and 2 - 4 corresponding to 'present'. Bacterial presence per colony was modeled using logit-linked binomial regression with the fixed effects of ∂15N, habitat, and DNA concentration (as a proxy for size), treating host genus as a random effect. Quantitative PCR estimates of 16S abundance per individual were modeled with a linear mixed model using the same fixed effects, but using colony nested within host genus as random effects. We checked for phylogenetic correlation of per-genus mean normalized bacterial abundance using Pagel's lambda (Pagel 1999) and Blomberg's K (Blomberg *et al.* 2003) with the R package phytools (Revell 2012), using the phylogeny from Moreau *et al.* (2013). For both GLMM and LMMs, Akaike Information Criteria were use to select the best model.

Results

Collections

From our primary field site at CICRA, we collected data for a total of 97 colonies, representing 97 morphospecies from 29 genera. Of these, 54 were collected from arboreal and 38 from terrestrial habitats. Voucher specimens for each colony have been deposited with the Centre de Ecologia y Biodiversidad (CEBIO) in Lima, Peru and the Museum of Comparative Zoology (MCZ) in Cambridge, MA, USA. Detailed collections information can be found in Table A3.1.

Visual microscopy survey

We were surprised to find that most ant guts surveyed by SybrGreen fluorescence microscopy did not harbor identifiable bacterial cells (N = 59; Fig. 3.1a). In these guts, although host nuclei were clearly stained and highly fluorescent (Fig. A3.2a), and gut contents could be seen spilling from the punctured gut under light microscopy and occasionally via autofluorescence (Fig. A3.2b), there were no visible DNA-containing cellular structures in the size range typical of bacteria. Several of the dissected guts did contain just a few apparent bacterial cells (visual rubric score of 1, N = 8). All individuals examined from the abundant and typically ground-nesting genera *Solenopsis* and *Pheidole* fell into these categories, as did all of the leaf-cutting ants, most individuals from ground-dwelling genera formerly grouped in the subfamily Ponerinae (including *Ectatomma* and *Pachycondyla*), and most of the individuals from the arboreal genera *Azteca*, *Crematogaster*, and *Pseudomyrmex*.

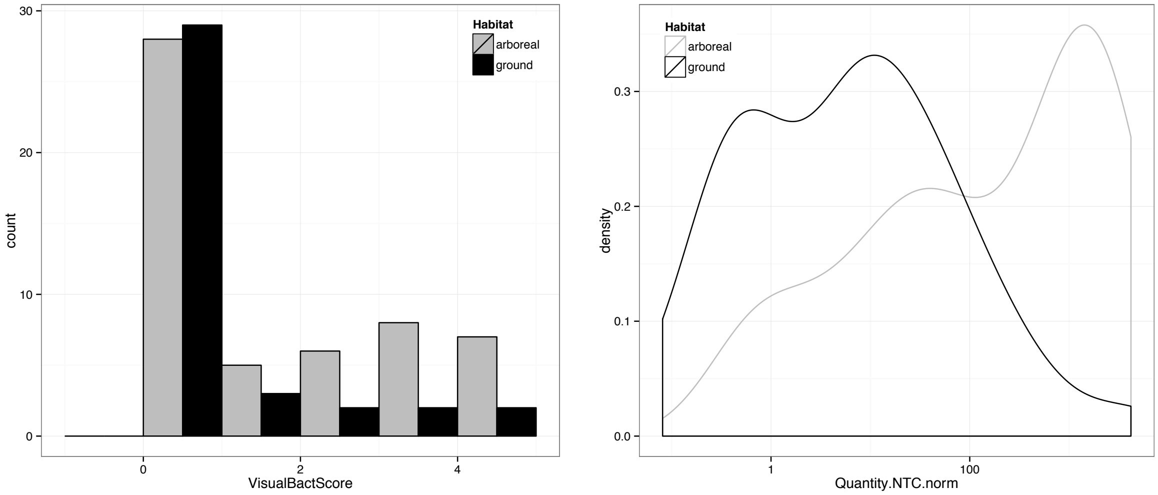


Fig. 3.1: a) Histogram of visual bacterial abundance estimates from *in situ* fluorescence microscopy. Estimates followed a roughly logarithmic scale (0 = no visible bacterial cells, 1 = tens, 2 = hundreds, 3 = thousands, 4 = tens of thousands; see Fig. S1). b) Kernel density plot of normalized bacterial abundance estimates from quantitative PCR.

By contrast, the density of bacterial cells in other ant guts was striking. Cell densities in guts of the abundant arboreal taxa *Camponotus*, *Cephalotes*, and *Dolichoderus*, were often so high that out-of-plane fluorescence inhibited photography using our field microscopy equipment (Fig. A3.2c). Moderate to high bacterial cell densities were frequently observed in army ants, including the ecitonine genera *Labidus* and *Eciton* as well as the cerapachyne genus *Acanthostichus* (visual scores 2-4, N = 4), although these genera also frequently appeared devoid of bacteria (visual scores 0-1, N = 4). The single individuals examined of myrmicine genera *Basiceros* and *Daceton* both hosted fairly high densities of putative bacterial cells. Individuals from genera that typically hosted few or no apparent bacterial cells (such as *Azteca* and *Crematogaster*) did occasionally contain high densities, although the reason for this variability is unclear.

At least three genera appeared to harbor bacterial cells in bacteriocytes localized to the midgut. These specialized cells were very clearly visible in the individuals we examined from *Camponotus*, appearing in the SybrGreen gut squash preps as networks of bright green patches intercalated with midgut cells (Fig. A3.2e,f). The process of puncturing the gut always disrupted a number of these bacteriocytes, spilling large numbers of the intracellular bacteria into the surrounding mounting medium. These cells were morphologically distinct, often quite large, and sometimes showed very long cells with intracellular DNA aggregation under high magnification suggestive of polyploidy (Fig. A3.2d; Fig. 3.2).

We also observed morphologically similar host cells, or putative bacteriocytes, in the midguts of individuals from one of the *Myrmelachista* morphospecies we examined (Fig. A3.2g,h). As in *Camponotus*, these appeared as fairly distinct bright green patches distributed around the midgut. The putative bacterial cells in these individuals presented as relatively large rods, though without the obviously anomalous morphologies frequently observed in *Camponotus* bacteria.

Individuals of many *Dolichoderus* also exhibited patterns of DNA fluorescence staining consistent with intracellular bacteria localized to the midgut. Unlike in *Camponotus*, where bacteriocytes were visible as clearly bounded cells, *Dolichoderus* midguts stained with SybrGreen appeared shrouded in a uniform green glow, largely obscuring the distinct host nuclei typically visible in other midguts. At higher magnification, these could be resolved as masses of morphologically unusual cells. Like the *Blochmannia* bacteria we observed erupting from *Camponotus* bacteriocytes, the putative intracellular bacteria in *Dolichoderus* were relatively large (Fig. 3.2). They were also often branched, again consistent with deficiencies in cell division and cell wall synthesis observed in other intracellular bacteria of insects (McCutcheon & Moran 2011). Along with these highly unusual cells, some *Dolichoderus* specimens exhibited high densities of smaller, coccoid bacterial cells. In at least one specimen for which we separately dissected midgut and hindgut compartments, these cells appeared localized to the hindgut.

Targeted microcopy using fluorescently-labeled universal bacterial 16S probes (FISH microscopy) supported our inferences from field-based SybrGreen microscopy. Whole-mount and resin sectioned guts from *Azteca*, *Pseudomyrmex*, and *Crematogaster* showed no evidence of bacteria, while several specimens from the army ants *Labidus* and *Eciton* showed small populations of bacterial cells localized to the walls of the hindgut (Fig. 3.2). By contrast, very large populations of bacteria could be readily seen in sections from *Camponotus*, *Cephalotes*, and *Dolichoderus*. In specimens of *Camponotus japonicus* (not collected from Peru, but used as representative sample from this well-studied genus), *Blochmannia* bacteria can be clearly observed in bacteriocytes interspersed among the midgut epithelia. In *Dolichoderus*, bacteria are visible forming a relatively uniform layer among cells in the outer portion of the midgut, as well as forming a dense mass in the pylorus and upper part of the ileum. In *Cephalotes*, bacterial cells form an aggregate that almost entirely fills the enlarged and highly folded ileum, as has been described previously through visible light and electron microscopy (Bution *et al.* 2007). Uniquely in *Cephalotes*, we also observed fluorescence indicating masses of bacterial cells in the distal part of the midgut lumen.

It should be noted that, for all ant specimens examined, high levels of tissue autofluorescence interfered with the relatively weak signal from monolabeled FISH probes. Autofluorescence was present in all channels, but especially pronounced in the green channel. Tissues lined with chitin (such as the crop and the rectum) displayed comparatively elevated levels of autofluorescence at longer wavelengths. Fat bodies and malphigian tubules also showed especially strong autofluorescence. Consequently, special care should be taken when interpreting FISH hybridizations from ant guts.

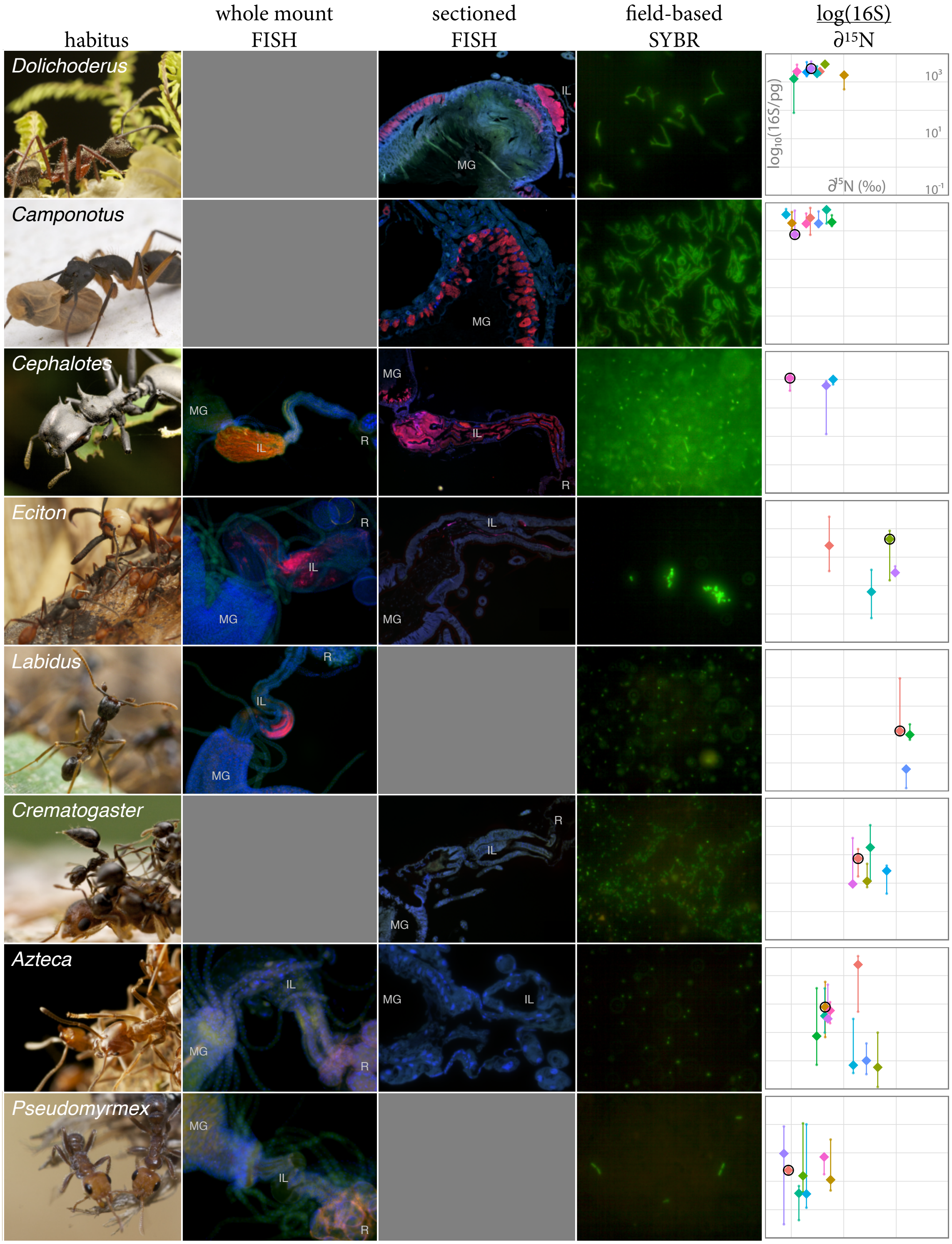

habitus
whole mount FISH
sectioned FISH
field-based SYBR
log(16S)
∂15N
Dolichoderus
IL
MG
log10(16S/pg)
∂15N (‰)
10^3
10^1
10^-1
Camponotus
MG
Cephalotes
MG
IL
R
MG
IL
R
Eciton
R
IL
MG
IL
MG
Labidus
R
IL
MG
Crematogaster
R
IL
MG
Azteca
IL
MG
R
MG
IL
Pseudomyrmex
MG
IL
R

Figure 3.2: Summarized microscopic and molecular evidence of gut bacterial abundance in eight common Peruvian ant genera. Column 1: Photographs and genus names (© JGS). Column 2: False-color FISH micrograph of whole-mount dissected guts. Tissue autofluorescence is visible in all three channels (blue, green, and red). DNA is stained with DAPI in the blue channel, and the universal bacterial probe Eub338 is hybridized in the red channel. MG: midgut; IL: Ileum; R: rectum. (© PL). Column 3: False-color FISH micrographs of resin-embedded tissue sections. Note bacteria present both putatively intracellularly (midgut wall) and extracellularly (in lumen of ileum) in *Dolichoderus*, and only intracellularly in *Camponotus*. Colors and labels as in Column 2 (© PL). Column 4: SYBR green fluorescence micrographs of bacteria from gut squashes. All images are uncropped and taken under identical magnification (40X objective). *Dolichoderus* image taken of bacteria from hindgut lumen (© JGS). Column 5: normalized $\log_{10}$ bacterial 16S copy number by $\partial^{15}N$ isotope ratio. Large diamonds represent median values per colony, with small points representing individuals. Lines indicate range of values observed for each colony. Each graph is on the same scale. The colony from which the SYBR green micrograph from Column 4 was taken is indicated by a black circle.

Quantitative PCR

Estimation of bacterial abundance via quantitative PCR targeting 16S rRNA largely corroborated visual estimates from field-based SybrGreen microscopy. The per-colony median 16S concentration correlated well with visual abundance estimates (Spearman's $\rho = 0.44$, $p << 0.001$), especially after normalizing by DNA concentration, a proxy for quantity of extracted tissue (Spearman's $\rho = 0.53$, $p << 0.001$). Colonies with a visual abundance score of 0 or 1 had DNA-normalized 16S concentrations statistically indistinguishable from one another, but significantly lower than colonies with visual scores of 2-4 (Fig. A3.3).

Estimates of 16S copy number correlated strongly with DNA concentration (Fig. A3.4a). For many smaller-bodied ant species, this meant that 16S quantities were below the detection threshold set by background amplification, or about 85 copies per µL. The lower bounds of detection may also have been affected by nonspecific amplification of host 18S molecules, which our primer set amplified with much lower affinity than bacterial 16S. Thus, relative differences between high- and low-abundance samples are likely to be underestimated by this method.

Despite these limitations, qPCR estimates revealed dramatic differences in the median bacterial abundance in colonies of different ant genera (Fig. 3.3; Fig. A3.5). To convey a sense of relative bacterial abundance independent of host body size, we use [DNA]-normalized values, expressed in 16S copies per pg DNA (Fig. A3.4b). These ranged from a minimum of 0.081 copies/pg in one ponerine colony, in which the median individual concentration was below the limit of detection despite a relatively high DNA concentration, to a maximum of 5537 copies/pg in a colony of *Camponotus*. The extremes were not dramatic outliers: the first and third quartiles were separated by more than two orders of magnitude (1Q: 3.77, 3Q: 931 copies / pg). Consistent with field microscopy observations, most genera we sampled had very low normalized bacterial abundances: only 10 of the 29 had median 16S counts above 100 copies / pg.

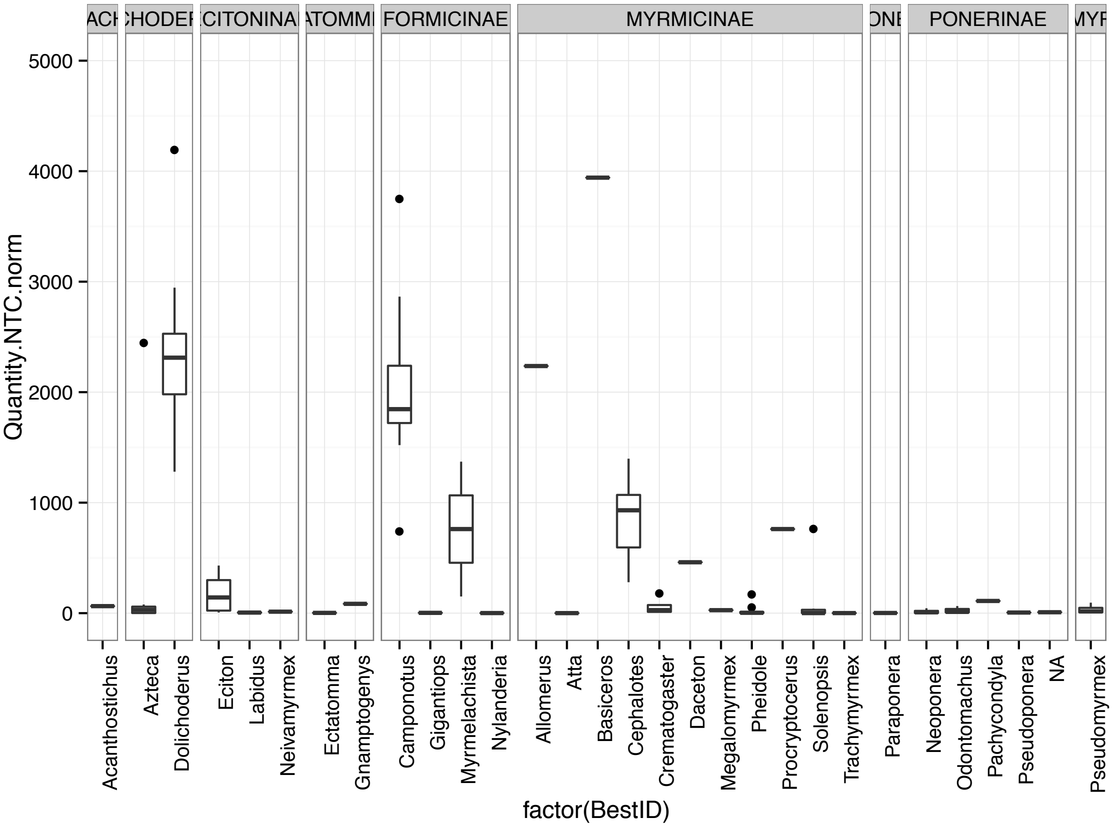


Figure 3.3: Normalized bacterial abundances by genus. Data shown are 16S qPCR counts, minus mean non-template control counts, divided by total DNA concentration. Each data point represents a single colony, taken as the median of three individuals.

Genera with high bacterial counts tended to be consistent among colonies. Colonies of the arboreal genera *Cephalotes*, *Camponotus*, and *Dolichoderus* all had consistently high normalized median 16S concentrations, with maximum and minimum values within an order of magnitude, despite relatively large numbers of colonies sampled. *Camponotus* and *Dolichoderus* colonies had somewhat higher normalized median 16S concentrations than did *Cephalotes* (medians per genera of 2031 and 2257 copies / pg vs 931.0 copies / pg, respectively). Of the twelve other genera for which we had sampled at least two colonies, only two – *Myrmelachista* and *Megalomyrmex* – had maximum normalized median concentrations within an order of magnitude of the minimum. As noted by field microscopy, both *Myrmelachista* colonies also hosted relatively high numbers of bacteria (151 and 1371 copies / pg), while both *Megalomyrmex* colonies we examined had fairly low numbers (26 and 27 copies / pg).

Correlation of bacterial abundance with ecological variables

To determine whether gut bacterial abundance correlates significantly with host ecology, we fit linear mixed models to visual and qPCR estimates of abundance, using host habitat (arboreal or terrestrial) and relative trophic position (inferred by $\partial^{15}N$ ratio) as fixed effects. Mean per-genus bacterial abundances were not significantly correlated across the ant phylogeny ($\lambda < 0.0001$, $p = 1$; $K = 0.54$, $p = 0.315$), allowing us to use genus and colony as random effects. As has been previously described from a similar sample of ants at a nearby site (Davidson *et al.* 2003), $\partial^{15}N$ ratios differed significantly by habitat (two-tailed students t test $p << 0.001$), with a mean of 6.57‰ in arboreal and 10.5‰ in terrestrial ants.

Fitting a generalized linear mixed model to presence or absence of bacteria in our field microscopy survey indicated that bacterial presence was significantly associated with habitat ($p = 0.0208$) as well as marginally by trophic position ($p = 0.0563$), but that the directionality of association with trophic position was opposite in arboreal and terrestrial habitats (interaction $p = 0.0165$). In arboreal ants, herbivorous colonies—those with lower $\partial^{15}N$ ratios—were more likely to host visible bacterial cells. But among terrestrial ants, the opposite was the case: bacteria tended to be found in more carnivorous ants (Fig. A3.7a). The model with this interaction term had a significantly better fit and lower Akaike Information Criterion values than models without it (Table A3.2).

Quantitative estimates of bacterial abundance via qPCR gave similar results. We fit linear mixed models of absolute bacterial 16S quantity with DNA concentration, habitat, and relative trophic position as fixed effects, taking colony nested within genus as random effects. As expected, 16S quantity correlated strongly with DNA concentration ($p < 0.0001$). Consistent with our findings above, bacterial abundances were higher in arboreal ants than in terrestrial ants ($p = 0.0084$) and correlated with relative trophic position ($p = 0.0081$), but the direction of correlation between bacterial abundance and relative trophic position differed in each habitat ($p = 0.0341$; Fig. A3.7b). As with the microscopy data, the model with an interaction term had significantly better fit and lower AIC (Table A3.3).

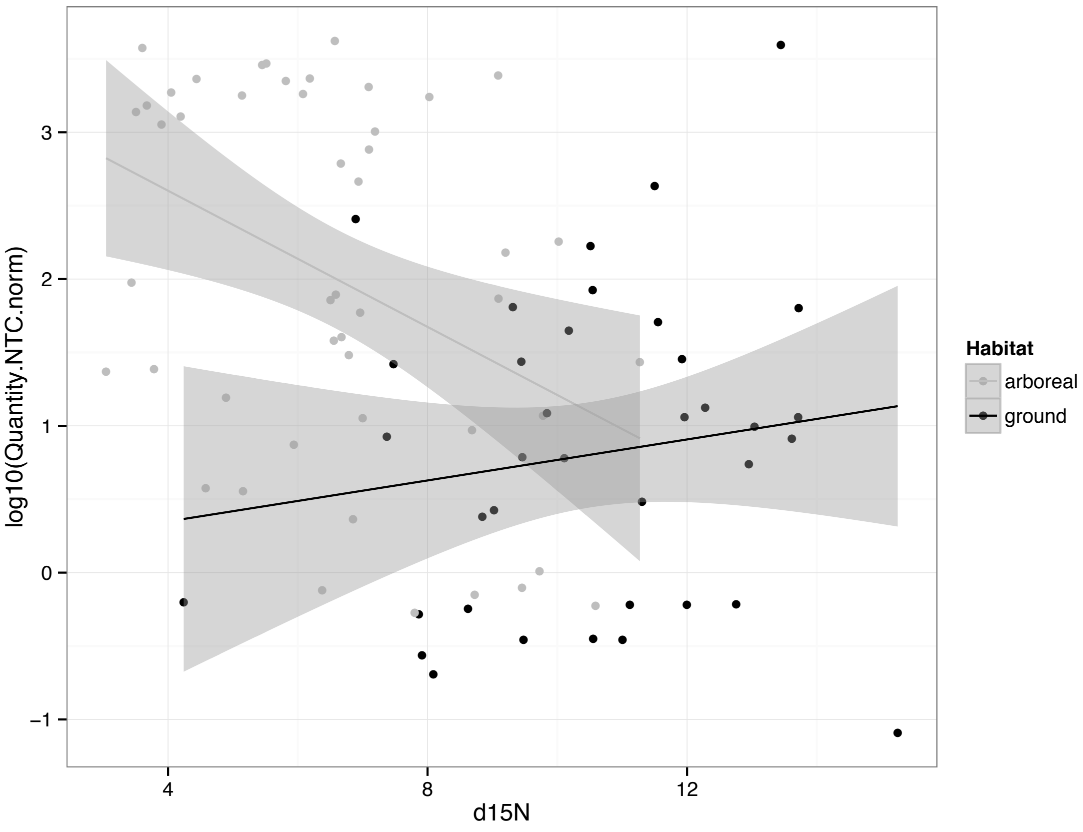


Figure 3.4: Normalized bacterial abundances (log qPCR 16S copy number per picogram DNA) by stable nitrogen isotope ratio. Each point represents the median value for a colony. Separate linear regressions (+/- 95% CI) fit to arboreal and ground-dwelling ants.

Discussion

Our findings support the hypothesis that symbioses with bacteria are systematically important to the dominance of ants in the tropical forest canopy (Davidson *et al.* 2003; Cook & Davidson 2006; Russell *et al.* 2009). Using two independent methods of characterizing bacterial abundance, we found bacteria to be both more abundant in arboreal ants, and a predictor of herbivory among arboreal ants. Surprisingly, most of the ants we surveyed had very few bacteria. We also identified abundant putative intracellular bacteria in two ant lineages for which such associations have not yet been described, *Myrmelachista* (Formicinae) and *Dolichoderus* (Dolichoderinae). Together, our findings present the beginnings of a systematic framework for understanding the relationship between diet and bacterial symbiosis in ants, and highlight the utility of absolute measures of abundance to interpreting the ecological role of bacterial symbioses.

Bacterial abundance and host ecology

Our findings support a relationship between bacteria and herbivory in canopy ants: almost all of the ants with very high concentrations of bacteria were canopy ants at the herbivorous extreme of the $\partial^{15}N$ scale, and the correlation between $\partial^{15}N$ isotope ratios and bacteria was significant for both microscopic and molecular measures of bacterial abundance. But while the ants with the highest concentrations of bacteria appeared to be mostly herbivorous, maintaining such high titers in worker guts does not seem to be essential to ant life in the canopy, or even to highly specialized herbivory. In our visual survey of ant guts, the distribution of bacterial abundance was strongly bimodal, with many arboreal individuals we surveyed not obviously hosting any bacterial cells at all (Fig. 3.1). Our qPCR-based estimates of bacterial abundance in arboreal ants were similarly bimodal. Ants represented by the lower peak of this distribution appear to be utilizing fundamentally different approaches to the challenge of acquiring nitrogen in the canopy.

The high-abundance peak of bacterial distribution was composed almost entirely of ants belonging to one of three taxa—*Camponotus*, *Cephalotes* (and its sister genus *Procryptocerus*), or *Dolichoderus*—that have previously been linked with bacterial symbioses. Of these, *Camponotus* symbioses are the best studied, with gamma-proteobacterial *Blochmannia* endosymbionts implicated in the recycling/upgrading of nitrogen from urea into essential amino acids (Feldhaar *et al.* 2007). The experimental evidence for a nutritional role in *Cephalotes* symbionts is to this point more limited (Jaffe *et al.* 2001). They host a moderately complex bacterial community in their gut lumen, comprising at least one species of Verrucomicrobia and several species alpha-, beta-, and gamma-proteobacteria. The *Cephalotes* gut community is both consistent and phylogenetically correlated across the genus (Sanders *et al.* 2013), and has shown some sensitivity to changes in diet (Hu *et al.* 2013). Little is known about the bacterial associates of *Dolichoderus* beyond a handful of 16S clones sequenced from a few individuals as part of other studies (Stoll *et al.* 2007; Russell *et al.* 2009; Anderson *et al.* 2012), but the sequence similarity of these clones to others sequenced from herbivorous ants has lead some to speculate that they play a similar functional role. Together, these three genera are responsible for virtually all of the correlation we observed between bacteria and herbivory: excluding them, there was no significant relationship between $\partial^{15}N$ isotope ratio and herbivory.

We posit that high bacterial abundances in these genera are necessary to sustain large nutrient fluxes. Despite major differences in the identity and physiology of their symbiotic associations, they have converged on a similar density. We measured median normalized 16S concentrations in these genera that were almost all within an order of magnitude of one another. Concentrations within *Cephalotes* were somewhat lower than in *Dolichoderus* and *Cephalotes*, though polyploidy in endosymbiotic bacteria and differences in per-genome 16S copy number make accurate extrapolation to absolute cell counts uncertain. What is certain is that workers from these three genera consistently maintain bacterial densities that are orders of magnitude greater than those found in most other ants. That such consistent associations have arisen independently in these three lineages, each with markedly herbivorous stable isotope signatures, lends additional credence to the hypothesis that bacteria play an important and convergent functional role in these canopy ants – and supports a connection between herbivory and the ant-specific lineage of *Bartonella* identified by Russell *et al.* in *Dolichoderus* and *Cephalotes*.

The much lower bacterial abundances we observed in almost all other arboreal ants suggests that they have evolved fundamentally different ecological and symbiotic strategies for life in the canopy. Some, like the abundant and ecologically dominant genera *Azteca* and *Crematogaster* (Wilson 1987), may simply be less herbivorous. These taxa rarely hosted any visible gut bacteria in our visual surveys, and had median normalized 16S concentrations two orders of magnitude lower than those of the high-abundance taxa. As with previous findings, they also had somewhat more omnivorous stable isotope profile—1 to 3‰ higher ∂^{15}N ratios than in *Camponotus*, *Cephalotes*, and *Dolichoderus*—suggesting that they complement their predominantly low-N liquid diets (Davidson *et al.* 2004) with moderate amounts of animal protein. Consistent with a strategy that pushes the boundaries of nitrogen availability, these taxa are reported to have among the lowest overall biomass nitrogen content and the highest behavioral preference for nitrogen-rich over carbohydrate-rich foods (Davidson 2005). Both of these genera typically have large, fast-growing colonies with presumably high overall demand for nitrogen.

More puzzling, perhaps, are the arboreal ants that harbored very low concentrations of bacteria, but still maintained depleted ∂^{15}N ratios in the same range as

Camponotus and *Dolichoderus*. Some of these may acquire their nitrogen from specialized associations with myrmecophytic plants. For example, *Neoponera luteola* is an obligate associate of *Cecropia pungara* (Yu & Davidson 1997), and the colony we measured had the most herbivorous isotope signature of any ant in our dataset (Fig. A3.8). The specialized food rewards provided by this species of *Cecropia* are especially nitrogen-rich for the genus (Folgarait & Davidson 1995), and may provide a major proportion of the ant's overall nitrogen budget. However, most of the *Pseudomyrmex* species we surveyed were not specialized residents of ant-plants, yet still had very low bacterial abundances and depleted ∂^{15}N ratios (the one obligate mutualist species we did survey, *P. triplaris*, was similar in both respects). Paradoxically, arboreal *Pseudomyrmex* have also been reported to have relatively limited behavioral preferences for nitrogen-rich foods compared to other arboreal ants (Davidson 2005) or to ground-nesting congeners (Dejean *et al.* 2014), suggesting that they have not evolved particularly strong behavioral imperatives for nitrogen acquisition. If the low densities we observed in the guts of arboreal *Pseudomyrmex* truly correspond to a limited role for bacteria in these ants' nitrogen economy, how should these foraging patterns be interpreted—as indications of adequate supply, or of limited demand? The relatively small colony size of free-living *Pseudomyrmex* species may simply require less nitrogen than the high-biomass colonies of *Azteca* and *Crematogaster*. Alternatively, arboreal *Pseudomyrmex* might form microbial associations at other lifestages (e.g. in the larval gut), or rely on alternative nitrogenous food sources, as in recent reports of fungal cultivation and consumption in the genus (Blatrix *et al.* 2012).

The comparative paucity of high bacterial loads among ground-nesting ants further supports the hypothesis that the extremely dense bacterial associations of some arboreal ants are adaptations particular to life in the canopy. In stark contrast to our findings for arboreal ants, ground-nesting ants showed no significant correlation between $\partial^{15}N$ isotope ratios and bacterial abundance—in fact, for visual estimates of abundance, there was a marginally significant trend towards higher densities in more carnivorous ants. Consistent with this trend, outside of *Camponotus*, *Dolichoderus*, and *Cephalotes*, we only definitively observed gut bacteria in FISH micrographs of the exclusively carnivorous army ants (Fig. 3.2).

Could bacterial associations facilitate extreme carnivory on the forest floor analogously to how they appear to have facilitated extreme herbivory in the canopy? Sequence-based surveys of bacteria have revealed consistencies among army ant microbiota (Funaro *et al.* 2011; Anderson *et al.* 2012) that seem to contrast with the highly variable communities that have been recovered from more generalist arboreal (Sanders *et al.* 2013) and terrestrial species (Lee *et al.* 2008; Meyer *et al.* 2011). High mortality in ants restricted to protein-rich foods—irrespective of carbohyrate content—also suggests a potential role for bacteria in ameliorating deleterious effects of obligate carnivory (Dussutour & Simpson 2012). If such associations do result in carnivorous ants hosting higher overall quantities of bacteria compared to more omnivorous species, the physiological demands of the association would appear to be satisfied by cell densities that are still orders of magnitude lower than in the canonical canopy-dwelling herbivores. More targeted investigations, using techniques with finer sensitivity at very low abundances, will be required to resolve this question.

Evidence for intracellular bacteria in multiple ant lineages

Nutritive intracellular endosymbionts are common in a great variety of insects (Moran *et al.* 2008), but, with the significant exception of *Blochmannia* endosymbionts in the speciose genus *Camponotus*, surprisingly absent among ants. After the initial description of intracellular bacteria in *Camponotus* and some species of *Formica*

(Blochmann 1888) over one hundred years ago, similar associations have not to our knowledge been described in any other ant lineages. We found microscopic evidence suggestive of gut-localized intracellular bacteria in two other arboreal ant lineages, suggesting that these associations may be considerably more widespread in ants than was previously thought.

Putative bacteriocytes in one of the two colonies of *Myrmelachista* we examined (colony JSC-108) appeared similar to those of *Camponotus*, and like in *Camponotus*, contained large, rod-shaped bacteria. *Myrmelachista* are specialized twig-nesters and frequent inhabitants of ant-plants, which form specialized structures to house and sometimes feed the ant inhabitants. Relatively little is known about the ecology of most species in the genus (Longino 2006), though the association between *M. schumanni* and the ant plant *Duroia hirsuta* results in dense, almost agricultural stands of the host plant due to pruning activity of the ants (Frederickson *et al.* 2005; Frederickson & Gordon 2007). A study of another plant associate, *M. flavocotea*, whose colonies nest in species of *Ocotea*, showed that workers of this species have a stable isotope signature much higher than that of their host plant, suggesting a substantial degree of carnivory (McNett *et al.* 2009). The two colonies in our dataset had sharply divergent $\partial^{15}N$ isotope ratios: colony JSC-137 (*M. schumanni*), which we recovered from the ant plant *Cordia nodosa*, was at about 9‰ similar to values reported from *M. flavocotea*. The colony in which we observed putative bacteriocytes in worker midguts (JSC-108) had a much more herbivorous signature; at 3‰, among the lowest values we recovered in our dataset. This colony also had a higher median 16S concentration by almost an order of magnitude (Fig. A3.8). While these observations are anecdotal, the correlated variation in presence of putative bacteriocytes, inferred diet, and overall bacterial abundance make *Myrmelachista* an attractive candidate for further study.

The putative intracellular bacteria we observed in *Dolichoderus* were even more striking. Individuals in this genus consistently harbored dense concentrations of bacteria in a blanket-like band around the proximate portion of the midgut (Fig. 3.2). These cells were quite large, with irregular, often heavily branched, morphologies. Branching morphologies have been reported in *Blochmannia* (Buchner 1965), and large, irregular phenotypes in other endosymbionts result from runaway gene loss associated with the bottlenecks of vertical transmission (McCutcheon & Moran 2011). While the data we present here cannot conclusively verify an intracellular location, the ultrastructural position of these cells (near the outer margin of midgut tissue, rather than interfacing with the gut lumen) and their derived morphology both strongly suggest it.

That *Dolichoderus* would share a symbiotic habit with *Camponotus* is perhaps unsurprising. In the forest habitats we sampled, the species are ecologically convergent, large-bodied ants, feeding largely on liquid exudates (Davidson *et al.* 2004). *Dolichoderus* specimens had DNA-normalized bacterial 16S concentrations and $\partial^{15}N$ ratios almost identical to those of *Camponotus*, among the most extreme of any ants we surveyed. Intriguingly, though, we observed several *Dolichoderus* colonies with bacterial cells that also appeared to be localized to the hindgut lumen, cells morphologically distinct from the large and highly branched cells of the midgut wall. These bacteria were especially apparent in sections of the hindgut stained with fluorescent probes, where they appear similar to the dense bacterial aggregates in the *Cephalotes* hindgut. While *Camponotus* seem to universally host intracellular *Blochmannia*, they are reputed, like some other insect hosts of endosymbionts, to be largely devoid of bacteria in the gut lumen (Feldhaar *et al.* 2007; Engel & Moran 2013). Whether these differences in symbiont localization have functional implications remains to be seen.

Most ants have very few bacteria

How unusual are the low bacterial densities we observed in ants? Direct numerical comparisons with organisms from other studies are challenging. Absolute bacterial abundances are only rarely reported in the literature (Engel & Moran 2013). When they are, the techniques used to derive them vary significantly, making direct comparison suspect. Furthermore, insects scale in body size across many orders of magnitude, so some normalization by host insect size is necessary. Given these caveats, the bacterial loads we measured in most ants were quite low compared to other insects. Normalized by estimated adult body weight, gut bacterial densities in low-abundance ants were on the order of 10^5 (*Ectatomma* and *Gigantiops*) to 10^6 (*Azteca* and *Crematogaster*) bacteria per gram, substantially lower than the ~10^8 estimated per gram in *Drosophila* (Ren *et al.* 2007; Engel & Moran 2013). By contrast, higher-abundance ants (*Cephalotes*, *Camponotus*, and *Dolichoderus*) had closer to 10^9 bacteria per gram—similar to values that have been estimated for aphids (Mira & Moran 2002), honey bees (Martinson *et al.* 2012), and humans (Savage 1977).

The shape of the bacterial abundance distribution within colonies of these low-abundance ant species hints at fundamental differences in the mechanisms underlying host/microbiome relationships among ant taxa (Fig. A3.9). High-abundance ant genera tended to have more normal distributions of bacterial abundance, implying that the loss of microbial cells through excretion and death is balanced in these taxa by cell birth and, potentially, ingestion. By contrast, distributions in low-abundance taxa were heavily right-skewed: while the median individual in low-abundance ant genera typically had very few detectable bacteria, we occasionally found individuals with much higher densities. These skewed distributions are reminiscent of similar patterns in *Drosophila*, which exhibit rapid decreases in bacterial abundance when starved or transitioned to sterile media, suggesting that the dynamics of the gut microbiome are weighted towards extinction (Broderick *et al.* 2014).

Understanding the significance of these infrequent, high-titer individuals will likely be important to understanding the nature of 'typical' ant-microbe interactions: do they represent dysbiotic individuals, in which host suppression of bacterial growth has failed? Do they reflect the recent ingestion of meals containing high concentrations of bacteria? Ants have evolved numerous ways of suppressing unwanted microbial growth inside their nests, including antibiotics derived endogenously from unique metapleural glands (Yek & Mueller 2010) and exogenously from specialized symbioses with actinomycete bacteria (Schoenian *et al.* 2011). This tendency towards microbial fastidiousness may extend to the inside of their guts, as well.

Practical considerations of low bacterial abundance

Our findings are especially important given the increasing ease and decreasing cost of amplicon-based microbial diversity profiling (Caporaso *et al.* 2011). These techniques describe relative, not absolute, differences in bacterial abundance. Consequently, comparisons may be easily made between samples with little or no awareness as to how they differ with respect to the total number of bacteria present – a variable that is likely to be profoundly relevant to biological interpretation. Such problems are compounded by the likely presence of contaminant amplicons derived from reagents, the effect of which itself varies proportionally to original concentration of bacteria in the sample (Salter *et al.* 2014).

These challenges are likely to be especially relevant in small-bodied insects with variable bacterial populations, like many of the ant genera we observed in this study. Even without considering variance in relative bacterial densities, ants from the same colony can span orders of magnitude in body size, leading to large differences in template quantity when amplifying from individuals. In practice, we have observed that within-colony variance community similarity is especially high in genera described here as having low overall bacterial abundances (Lee *et al.* 2008; Meyer *et al.* 2011; Sanders *et al.* 2013). These samples may have been affected by undetected reagent contamination, or simply reflect higher intrinsic variation in communities with fewer or no endogenous commensals. Pairing amplicon-based community profiling data with direct estimates of absolute bacterial abundance will help to provide important biological context to sequence diversity surveys.

Conclusion

The explosion of 16S sequencing studies has justifiably led to an explosion of interest in animal microbiota (McFall-Ngai *et al.* 2013). Sequencing gives easy access to information about the composition of microbial communities, leading to extraordinary insights into the function and diversity of host-associated bacteria. Here, by demonstrating that gut bacterial densities help to explain the relationship between diet and habitat in rainforest ants, we have shown that simply surveying the abundance of these microbes can be useful as well, providing insights into ecology and potential function not possible simply by sequencing.

We humans host about a kilogram of bacteria in our gut ("Microbiology by numbers." 2011); a *Cephalotes* ant, scaled to human size, would harbor roughly the same amount. The bacteria in the gut of *Gigantiops destructor*, similarly scaled, would weigh about as much as a roast coffee bean. We propose that these differences in magnitude correspond to differences in physiology with major relevance to the host.

Acknowledgements

We thank Frank Asorza and Stefan Cover for assistance with identification of specimens; Gabriel Miller, Lina Arcila Hernandez, Antonio Coral, and the staff of CICRA for assistance with collections and field work; Stephen Worthington and Kareem Carr for statistical advice; and Jacob Russell and Peter Girguis for providing laboratory facilities and support. Funding for this work was provided in part by a Putnam Expedition grant and NSF DDIG to JGS.

APPENDIX 1

Chapter 1 Supplemental

Noise reduction and clustering

Clustering at 99% similarity is likely to be especially subject to "noise" generated by polymerase error and sequencer error. These errors dramatically increase the number of estimated OTUs (Kunin *et al.* 2010; Quince *et al.* 2011), and we would expect these spurious OTUs to obscure patterns of differentiation among samples, especially for binary beta diversity metrics (such as Jaccard and Sorensen) that are especially sensitive to rare species. In our *Cephalotes* data, we have attempted to correct for these effects using AmpliconNoise (Quince *et al.* 2011). Because AmpliconNoise attempts to model and correct for polymerase error (which may itself mirror natural variation in the population) as well as sequencing error, it may err in the opposite direction, removing "true" sequences that are present in samples at low abundance (Gaspar & Thomas 2013). This will in turn suppress the effect of rare taxa on binary beta diversity metrics. We analyzed the effect of denoising on our dataset using the methods suggested by Gaspar *et al.* (2013), and found that approximately 0.53% of bases were modified in the process (Table A1.4), particularly as a result of the "accordion effect" noted by those authors. This error rate is close to the absolute raw error rate observed for 454 sequencing of mock communities (Quince *et al.* 2011).

Despite imperfections in available denoising techniques, two lines of evidence suggest to us that use of AmpliconNoise-treated 99% OTUs is appropriate for our dataset. First, in tests of the AmpliconNoise algorithm using data from known communities of similar diversity to what we estimated for the *Cephalotes* gut, masked 'real' variation was low even at 99% clustering, and much lower than observed for other noise reduction approaches (Quince *et al.* 2011). Second, while *Cephalotes* clades clustered most frequently using 99% OTUs, the trend towards greater phylogenetic correlation at narrower OTU definitions was also apparent from 93 to 97% in both Mantel tests (Fig. 1.5), network clustering (Fig. 1.4 and A1.5), and PCoA plots (Fig. A1.6), suggesting that the greater correlation observed using 99% clustering is not solely an artifact of that single parameter choice.

Alpha diversity: microbiome richness and taxonomic composition

By several measures of alpha diversity, including observed species richness, Shannon diversity, and the Chao1 richness estimator, the *Cephalotes* gut samples were fairly simple at the 97% OTU level (Table A1.3). When rarified to a constant depth of 1000 reads, these samples were comprised on average of just 20 unique OTUs, with a Chao1 estimated total richness of 28 unique OTUs per individual. Samples sequenced from whole *Cephalotes* gasters yielded somewhat greater richness, with 34 observed and 50 predicted OTUs. Samples from non-Cephalotine guts were considerably more diverse, with *Azteca* guts yielding 51/56 and *Pseudomyrmex* guts yielding 84/102 observed/predicted OTUs (Fig. A1.2; none of the *Crematogaster* samples yielded more than 1000 sequences after filtering, and are not shown). Clustering algorithm made only modest differences in alpha diversity measurements (Table A1.3); unless otherwise stated, values presented were clustered using UCLUST.

Cephalotes gut microbiomes were dominated by a consistent mix of Proteobacteria and Verrucomicrobes (Fig. 1.2), as has been reported previously for this genus (Russell *et al.* 2009; Anderson *et al.* 2012). In our dataset, a single lineage of Verrucomicrobe assigned to the class Opitutales comprised the vast majority of reads, with 71% of reads in the average *Cephalotes* gut sample. The remaining 29% of reads from *Cephalotes* guts were divided among several lineages of β-Proteobacteria (class Burkholderiales, 10%), γ-Proteobacteria (Xanthomonadales [2.6%], and an unassigned γ-Proteobacterial lineage [5.5%]), and α-Proteobacteria (class Rhizobiales, 2.6%), corresponding to bacterial lineages previously characterized as 'ant-specific' by Russell *et al.* (2009). Reads assigned to the Bacteroidetes appeared at low abundance (averaging 1.1%) in about 1/3 of *Cephalotes* gut samples, though they were substantially more abundant (ranging from 5-15% of total relative abundance) in three individuals sampled from one of the two colonies of *C. atratus*. Seven to eight 'core' OTUs were recovered in 50% or more of the *Cephalotes* samples, depending on OTU width. These fell into roughly eight phylogenetic clades in a maximum-likelihood tree of *Cepahlotes*-specific sequences from Anderson *et al.* (2012), corresponding to clade Opitutales A-1, Xanthomonadales A-2, Pseudomonadales A-3, and Burkholderiales A2, B, and C, and Rhizobiales A2 (Fig. A1.3).

The relative proportional representation of each of these bacteria remained fairly constant across *Cephalotes* samples, with a few notable exceptions. Most visibly, samples derived from ethanol-preserved gasters were substantially reduced in relative representation of Verrucomicrobes (20% ethanol gasters vs. 71% for dissected guts). This held true both when comparing between museum specimens and the freshly collected sample set, and for the one colony of *C. pusillus* for which we sequenced both from dissected guts and from gasters that had simultaneously been preserved in ethanol (25% ethanol gasters vs 94% guts). It is uncertain whether this discrepancy is due to differences in preservation and extraction efficiency or to the additional tissue present in the gaster samples. Additional studies with more absolute quantitative fidelity, perhaps using bacterial-specific fluorescent microscopy or targeted qPCR, will be necessary to resolve this question.

In another intriguing departure from the norm, all three individuals from one colony of *C. persimilis* had substantial representation of an ε-Proteobacterium assigned to the Campylobacteraceae. For two of these individuals, the ε-Proteobacterium comprised more than 82% of all reads. Although it didn't exceed 10% of any other sample, and only exceeded 1% of reads in two others, 36 *Cephalotes* samples had at least one sequence assigned to this ε-Proteobacterium. Reads assigned to the Campylobacteriales also dominated at least one *C. varians* individual from a previous study (Kautz *et al.* 2012). This OTU may represent a commensal lineage that is opportunistically pathogenic, or one that dominates an alternative ecological state -- i.e. an alternative 'enterotype' -- reflective of some altered aspect of the gut environment.

The two gut communities we sequenced from *Procryptocerus*, the sister genus to *Cephalotes*, showed taxonomic profiles fairly similar to *Cephalotes*, with abundant representation of Verrucomicrobes and Proteobacteria. These two samples came from ethanol-preserved gasters, and as with the ethanol-preserved *Cephalotes* gasters, showed a lower proportion of Verrucomicrobes than did most *Cephalotes* gut samples (8% and 56%).

Microbiomes from *Pseudomyrmex*, *Crematogaster*, and *Azteca* guts showed very little taxonomic overlap with the cephalotine samples. A plurality of reads in each of these genera matched closely with known insect pathogens. Sixty percent of cumulative reads from the three *Pseudomyrmex* samples were 99% identical to 16S from the Group I spiroplasmas, many of which are pathogenic in insects and plants (Regassa & Gasparich 2006). Reads with >95% homology to a wide range of *Wolbachia* endosymbionts/parasites comprised 41% and 71% of *Crematogaster* and *Azteca* samples, respectively. The remainder of reads in these samples were distributed among a number of bacterial taxa, including lineages of Clostridia and Bacteroidia commonly associated with vertebrate guts, Bacilli in the Staphylococcaceae, and γ-Proteobacteria in the Pseudomonadales, Enterobacteriales, and Oceanospirillales.

Beta diversity: technical variance

Remarkably, similarities between gut communities from nestmates were statistically indistinguishable from similarities between replicate PCRs from the same individual (Fig.1.3). This unintuitive outcome is likely to have several causes, in addition to the small sample size [n=3]. First, samples were amplified by the sequencing facility in singlicate and at fairly high cycle number (35), which is likely to have contributed to random PCR bias and exaggerated inter-sample and inter-replicate differences. Second, the three technically replicated samples happened to have very low alpha diversity and were overwhelmingly dominated by a single Verrucomicrobial OTU. This leaves very few of the remaining reads to sample the rest of the diversity, increasing the chances that lower-abundance OTUs were 'missed' upon resequencing. Since the Jaccard metric doesn't take abundance into account, it is particularly sensitive to these rare OTUs. Finally, technical replicates were sequenced on a separate flow cell, so that any effect of between-flow-cell variance would be observed in between-replicate but not between-nestmate comparisons. While larger sample sets with greater statistical power would likely reveal significant within-colony community variation, the homogenizing effect of the colony is also apparent in our finding that *Cephalotes* nestmates have microbiomes significantly more similar to one another than do conspecific non-nestmates.

16S analysis script

We used the following Unix/Bash shell script to automate analysis; the script is provided as reference for those wishing to replicate this analysis, or perform similar analyses on their own datasets. The script is a modified version of the tutorial analysis script provided in QIIME v1.4.0. All scripts are also provided in the Dryad data repository under DOI doi:10.5061/dryad.023s6.

```
#!/bin/bash

mkdir $1
cd $1
pct_id=.$1

seqs_file=seqs.fna
map_file=cephalotes_denoise_map_tree.txt
```

```
gg_db_dir=gg_otus_4feb2011
align_temp_file=~/qiime/data/core_set_aligned.fasta.imputed
lanemask_file=~/qiime/data/lanemask_in_1s_and_0s
rarify=1000
parameters_file=final_parameters.txt
beta_metrics='binary_dist_jaccard binary_dist_sorensen_dice dist_bray_curtis \
unweighted_unifrac weighted_unifrac'

#Pick OTUs and remove chimeras with CD-HIT.
pick_otus.py -i $seqs_file -s $pct_id -o cdhit_results/

#Pick rep set
pick_rep_set.py -i cdhit_results/*otus.txt -f $seqs_file \
-o rep_set_raw.fna

#Chimera picking with UCHIME

#Script to add abundance information to rep set fasta, per UCHIME input
requirements
add_abundances.py cdhit_results/*otus.txt rep_set_raw.fna uchime/
rep_set_abund.fna

#De Novo chimera detection
usearch -uchime uchime/rep_set_abund.fna -chimeras uchime/de_novo_ch.fasta \
 -nonchimeras uchime/de_novo_good.fasta -uchimeout uchime/de_novo_results.uch \
 -uchimealns uchime/de_novo_results.alns -log uchime/de_novo_log.txt

#Reference chimera detection
usearch -uchime uchime/rep_set_abund.fna -chimeras uchime/ref_ch.fasta \
 -nonchimeras uchime/ref_good.fasta -rev -uchimeout uchime/ref_results.uch \
 -uchimealns uchime/ref_results.alns -log uchime/ref_log.txt \
 -db gold.fa

#Script to generate intersection of chimeric sequence IDs
intersect_fasta_ids.py uchime/ref_ch.fasta uchime/de_novo_ch.fasta uchime/
intersect.txt uchime/union.txt

#Remove chimeric seqs from rep set fasta
filter_fasta.py -f rep_set_raw.fna -o rep_set.fna -s uchime/union.txt -n

#assign taxonomy using retrained greengenes database
assign_taxonomy.py -i rep_set.fna \
-t $gg_db_dir/taxonomies/greengenes_tax_rdp_train.txt \
-r $gg_db_dir/rep_set/gg_97_otus_4feb2011.fasta -o rdp_assigned_taxonomy_gg/

#align rep set sequences
align_seqs.py -i rep_set.fna -t $align_temp_file -o pynast_aligned
filter_alignment.py -i pynast_aligned/rep_set_aligned.fasta \
-m $lanemask_file -o filtered_alignment/

#make OTU tree
make_phylogeny.py -i filtered_alignment/rep_set_aligned_pfiltered.fasta \
-o filtered_alignment/rep_set.tre

#make otu table
mkdir otu_table
make_otu_table.py -i cdhit_results/*otus.txt \
-t rdp_assigned_taxonomy_gg/rep_set_tax_assignments.txt \
-o otu_table/unsorted_prefiltered_otu_table.txt

#sort otu table
sort_otu_table.py -i otu_table/unsorted_prefiltered_otu_table.txt \
-o otu_table/prefiltered_otu_table.txt -m $map_file -s ColonySorter
```

```
#get rid of poor samples from the one run
filter_by_metadata.py -i otu_table/prefiltered_otu_table.txt \
-m $map_file -s 'Exclude:0' -o otu_table/otu_table.txt

#single rarifaction and recombination
#This uses a modified version of the single_rarefaction.py script in QIIME.
#This script is modified to retain samples with lower sequence counts.
single_rarefaction_include_small.py -i otu_table/otu_table.txt \
-o otu_table/otu_table_rarified.txt -d $rarify --small_included

#summarize by individual column in mapping file, also by colony and by species
summarize_otu_by_cat.py -i $map_file -c otu_table/otu_table_rarified.txt \
-m HostSpecies -o otu_table/otu_table_rarified_HostSpecies.txt
summarize_otu_by_cat.py -i $map_file -c otu_table/otu_table_rarified.txt \
-m SpeciesColony -o otu_table/otu_table_rarified_SpeciesColony.txt

#OTU Heatmap
echo "OTU Heatmap"
make_otu_heatmap_html.py -i otu_table/otu_table_rarified.txt \
-t filtered_alignment/rep_set.tre -o otu_table/OTU_Heatmap_rarified
make_otu_heatmap_html.py -i otu_table/otu_table_rarified_HostSpecies.txt \
-t filtered_alignment/rep_set.tre \
-o otu_table/OTU_Heatmap_rarified_HostSpecies
make_otu_heatmap_html.py -i otu_table/otu_table_rarified_SpeciesColony.txt \
-t filtered_alignment/rep_set.tre -o otu_table/OTU_Heatmap_rarified_TreeName

#OTU Network
echo "OTU Network"
make_otu_network.py -m $map_file -i otu_table/otu_table_rarified.txt \
-o otu_table/OTU_Network

#Make Taxa Summary Charts
echo "Summarize taxa"
summarize_taxa_through_plots.py -i otu_table/otu_table_rarified.txt \
-o taxa_summary -m $map_file

#Alpha diversity
echo "Alpha rarefaction"
alpha_rarefaction.py -i otu_table/otu_table.txt -m $map_file \
-o alpha/ -p $parameters_file -t filtered_alignment/rep_set.tre

#Beta diversity
echo "Beta diversity and plots: all"
beta_diversity_through_plots.py -i otu_table/otu_table.txt \
-m $map_file -o beta/ -t filtered_alignment/rep_set.tre \
-p $parameters_file
echo "Beta diversity and plots: rarified"
beta_diversity_through_plots.py -i otu_table/otu_table_rarified.txt \
-m $map_file -o beta${rarify}/ -t filtered_alignment/rep_set.tre \
-p $parameters_file

echo "Jackknifed beta diversity"
jackknifed_beta_diversity.py -i otu_table/otu_table.txt \
-t filtered_alignment/rep_set.tre -m $map_file -p $parameters_file \
-o jack -e 700

echo "Make Bootstrapped Tree"
for metric in $beta_metrics
do
echo $metric
make_bootstrapped_tree.py -m jack/$metric/upgma_cmp/master_tree.tre \
-s jack/$metric/upgma_cmp/jackknife_support.txt \
-o jack/$metric/upgma_cmp/jackknife_named_nodes.pdf
done
```

Beta-diversity community clustering sensitivity

To automate the comparison of jackknifed community clustering data against the host phylogeny, we used the following Unix/Bash shell script. This script iterates over a series of results directories clustered at different OTU widths ('93 95 97 99') generated using the above base analysis script. It then iterates over a series of jackknifed community clustering results for different beta diversity measures ('binary_dist_jaccard binary_dist_sorensen_dice dist_bray_curtis unweighted_unifrac weighted_unifrac') and compares them against an input host phylogeny ('host_tree.tre') using the QIIME tree_compare.py script. All scripts are also provided in the Dryad data repository under DOI doi:10.5061/dryad.023s6.

```
#!/bin/bash

base_dir='./base_analysis'
otus='93 95 97 99'
beta_metrics='binary_dist_jaccard binary_dist_sorensen_dice dist_bray_curtis\
unweighted_unifrac weighted_unifrac'
guide_tree='./host_tree.tre'

mkdir $base_dir/bdiv_summary

for metric in $beta_metrics
do
echo $metric
thismetric=$metric
for otu in $otus
do
echo $otu
echo $thismetric

tree_compare.py -m $guide_tree \
-s $base_dir/${otu}/jack/${metric}/rare_upgma \
-o $base_dir/bdiv_summary/${metric}_${otu}
done

done
```

Beta-diversity Monte Carlo tests

To assess whether measures of beta-diversity differed significantly between different subsets of our data, we employed Monte Carlo permutation tests. Because sets of pairwise distances violate assumptions of independence among observations critical to standard parametric and non-parametric significance tests, and because the hierarchical (individuals within colonies within species within genera) and unbalanced nature of our sample set precluded use of the simple within/between group Monte Carlo tests in QIIME or of the *permute* package in *vegan*, we implemented our own permutation schemes in R.

Essentially, our script decomposed a distance matrix into a vector of pairwise distances, grouped sets of pairwise distances by category (e.g. intra-colony distances of *Cephalotes*, intra-colony distances of non-*Cephalotes*, and other), and generated a test statistic (Student's T-statistic) for the difference between the two focal groups. It then randomly reassigned membership among the focal groups, while retaining within-strata relationships as appropriate to the specific test (see below). Significance was calculated as the proportion of permuted test statistics lower than the unpermuted test statistic, inclusive of the unpermuted value (one-tailed tests).

Actual scripts used are provided in the Dryad data repository under DOI doi: 10.5061/dryad.023s6. The following comparisons were tested, each with 1000 permutations:

Intra-colony vs replicate PCRs:

To place technical variance due to PCR and sequencing in the context of variance within a colony, pairwise distances for *Cephalotes pusillus* colony mates JS0102, JS0103, and JS0104 were separated into between-replicate distances (e.g. JS0102 vs JS0102a.1) and within-colony, within-replicate distances (e.g. JS0102 vs JS0103 or JS0102a.1 vs JS0103a.1). Group assignment was permuted.

Within-clade vs between-clade (Cephalotes inter-species):

To determine whether interspecific distances were significantly greater between species from different clades than between species from the same clade. Intra-colony and intra-specific distances were excluded, and clade assignment permuted among colonies.

Within-species vs between-species (Cephalotes inter-colony):

To determine whether interspecific distances were significantly greater than intraspecific (but between-colony) distances. Species assignment was permuted among colonies.

Within-colony vs between-colony (Cephalotes intraspecific):

To determine whether intraspecific distances were greater between colonies than within colonies for *Cephalotes*. Tests (and permutations) were restricted to those *Cephalotes* species for which we had data for more than one colony. Colony assignment of individuals was permuted, but restricted to within species.

Within-colony Cephalotes vs within-colony non-Cephalotes

To determine whether within-colony distances for non-*Cephalotes* nestmates were significantly greater than within-colony distances for *Cephalotes* nestmates. Only colonies with multiple individuals were included (i.e. excluding museum specimens). Genus assignment was permuted among colonies.

Intra-colony vs inter-species (non-cephalotines)

To determine whether distances between species for non-cephalotine outgroups were significantly greater than distances within colonies, colony assignment (and hence species and genus assignment, as there was only one colony per outgroup species/genus) was permuted among individuals.

Table A1.1: sample information

Colony	Clade	Species	Voucher†	Individuals	Source	Lat	Lon	Tissue
43	-	*Azteca* sp.	*JS0524*	3	Brazil	-19.11	-48.24	gut
31	-	*Pseudomyrmex* sp.	*JS0372*	3	Brazil	-19.11	-48.24	gut
35	-	*Crematogaster* sp.	*JS0411*	3	Brazil	-19.11	-48.24	gut
h	-	*Procryptocerus* sp. 1	*ER9*	1	Museum	10.35	-67.68	gaster
i	-	*Procryptocerus* sp. 2	*JS0687*	1	Peru	-12.50	-70.10	gaster
15	atratus	*Cephalotes atratus*	*JS0164*	3	Brazil	-19.11	-48.24	gut
46	″	″	*JS0557*	3	Peru	-12.50	-70.10	gut
e	″	*Cephalotes placidus*	*DK07/21*	1	Museum	1.92	-67.06	gaster
g	basalis	*Cephalotes mompox*	*ER4*	1	Museum	10.46	-67.77	gaster
55	texanus	*Cephalotes rohweri*	*JS0719*	3	Arizona	32.20	-110.90	gut
f	″	*Cephalotes scutulatus*	*DK30*	1	Museum	9.75	-83.75	gaster
50	umbraculatus	*Cephalotes umbraculatus*	*JS0603*	3	Peru	-12.50	-70.10	gut
11	clypeatus	*Cephalotes clypeatus*	*JS0089*	3	Brazil	-19.11	-48.24	gut
lmah1	″	″	-	3	Peru	-12.50	-70.10	gut
32	laminatus	*Cephalotes minutus*	*JS0387*	3	Brazil	-19.11	-48.24	gut
51	″	“	*JS0632*	3	Peru	-12.50	-70.10	gut
12	″	*Cephalotes pusillus*	*JS0111*	6*	Brazil	-19.11	-48.24	gaster/gut
14	″	″	*JS0141*	3	Brazil	-19.11	-48.24	gut
27	″	″	*JS0338*	3	Brazil	-19.11	-48.24	gut
45	″	*Cephalotes simillimus*	*JS0547*	3	Peru	-12.50	-70.10	gut
47	″	*Cephalotes spinosus*	*JS0571*	3	Peru	-12.50	-70.10	gut
53	angustus	*Cephalotes targionii*	*JS0672*	3	Peru	-12.50	-70.10	gut
a	fiebrigi	*Cephalotes bohlsi*	*AW0620*	1	Museum	-23.44	-58.44	gaster
25	pallens	*Cephalotes pallens*	*JS0297*	3	Brazil	-19.11	-48.24	gut
49	″	*Cephalotes pallidoides*	*JS0590*	3	Peru	-12.50	-70.10	gut
10	″	*Cephalotes pellans*	*JS0063*	3	Brazil	-19.11	-48.24	gut
19	pinelii	*Cephalotes maculatus*	*JS0223*	3	Brazil	-19.11	-48.24	gut
lmah2	″	″	-	3	Peru	-12.50	-70.10	gut
21	grandinosus	*Cephalotes grandinosus*	*JS0251*	3	Brazil	-19.11	-48.24	gut
18	″	*Cephalotes persimilis*	*JS0209*	3	Brazil	-19.11	-48.24	gut
23	″	″	*JS0270*	3	Brazil	-19.11	-48.24	gut
26	″	″	*JS0323*	3	Brazil	-19.11	-48.24	gut
d	crenaticeps	*Cephalotes crenaticeps*	*ER8*	1	Museum	10.41	-67.61	gaster
b	coffeae	*Cephalotes peruviensis*	*DK25*	1	Museum	-0.80	-75.51	gaster
c	″	*Cephalotes setulifur*	*DK27*	1	Museum	9.75	-83.75	gaster
24	depressus	*Cephalotes borgmeieri*	*JS0282*	3	Brazil	-19.11	-48.24	gut
48	″	*Cephalotes cordatus*	*JS0686*	3	Peru	-12.50	-70.10	gut
20	″	*Cephalotes eduarduli*	*JS0240*	3	Brazil	-19.11	-48.24	gut

* technical replicate sequencing performed on three of these individuals
† vouchers are deposited in the Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138.

Table A1.2: Number of sequences for each sample in this study.

SampleID	Individual	Colony	Species	# sequences
JS0516.1	JS0516	43	*Azteca sp.*	63
JS0517.1	JS0517	43	*Azteca sp.*	3212
JS0518.1	JS0518	43	*Azteca sp.*	8895
JS0366	JS0366	31	*Pseudomyrmex sp.*	1283
JS0367	JS0367	31	*Pseudomyrmex sp.*	3642
JS0368	JS0368	31	*Pseudomyrmex sp.*	1388
JS0413	JS0413	35	*Crematogaster sp.*	731
JS0414	JS0414	35	*Crematogaster sp.*	240
JS0416	JS0416	35	*Crematogaster sp.*	947
JS0759	JS0759	h	*Procryptocerus sp. 1*	4705
JS0687	JS0687	i	*Procryptocerus sp. 2*	572
JS0151	JS0151	15	*Cephalotes atratus*	4137
JS0152	JS0152	15	*Cephalotes atratus*	3103
JS0153	JS0153	15	*Cephalotes atratus*	2748
JS0548.1	JS0548	46	*Cephalotes atratus*	974
JS0549.1	JS0549	46	*Cephalotes atratus*	724
JS0550.1	JS0550	46	*Cephalotes atratus*	462
JS0756	JS0756	e	*Cephalotes placidus*	2899
JS0758	JS0758	g	*Cephalotes mompox*	2774
JS0708	JS0708	55	*Cephalotes rohweri*	3194
JS0709	JS0709	55	*Cephalotes rohweri*	4694
JS0710	JS0710	55	*Cephalotes rohweri*	639
JS0757	JS0757	f	*Cephalotes scutulatus*	932
JS0591	JS0591	50	*Cephalotes umbraculatus*	2933
JS0592	JS0592	50	*Cephalotes umbraculatus*	2717
JS0593	JS0593	50	*Cephalotes umbraculatus*	4914
JS0068	JS0068	11	*Cephalotes clypeatus*	6748
JS0069	JS0069	11	*Cephalotes clypeatus*	2367
JS0070	JS0070	11	*Cephalotes clypeatus*	7391
JS0690	JS0690	Imah1	*Cephalotes clypeatus*	642
JS0691	JS0691	Imah1	*Cephalotes clypeatus*	478
JS0692	JS0692	Imah1	*Cephalotes clypeatus*	843
JS0375	JS0375	32	*Cephalotes minutus*	3689
JS0376	JS0376	32	*Cephalotes minutus*	814
JS0377	JS0377	32	*Cephalotes minutus*	4273
JS0623.1	JS0623	51	*Cephalotes minutus*	1691
JS0624.1	JS0624	51	*Cephalotes minutus*	2048
JS0625.1	JS0625	51	*Cephalotes minutus*	2120
JS0102	JS0102	12	*Cephalotes pusillus*	2630
JS0102a.1	JS0102	12	*Cephalotes pusillus*	2115
JS0103	JS0103	12	*Cephalotes pusillus*	2093

Table A1.2 (Continued): Number of sequences for each sample in this study.

SampleID	Individual	Colony	Species	# sequences
JS0103a.1	JS0103	12	*Cephalotes pusillus*	3781
JS0104	JS0104	12	*Cephalotes pusillus*	4442
JS0104a.1	JS0104	12	*Cephalotes pusillus*	2918
JS0111a.1	JS0111a	12	*Cephalotes pusillus*	303
JS0111b.1	JS0111b	12	*Cephalotes pusillus*	1127
JS0111d.1	JS0111d	12	*Cephalotes pusillus*	965
JS0124.1	JS0124	14	*Cephalotes pusillus*	2020
JS0125.1	JS0125	14	*Cephalotes pusillus*	661
JS0126.1	JS0126	14	*Cephalotes pusillus*	1917
JS0328.1	JS0328	27	*Cephalotes pusillus*	2514
JS0329.1	JS0329	27	*Cephalotes pusillus*	753
JS0330.1	JS0330	27	*Cephalotes pusillus*	2283
JS0535	JS0535	45	*Cephalotes simillimus*	3891
JS0536	JS0536	45	*Cephalotes simillimus*	2431
JS0537	JS0537	45	*Cephalotes simillimus*	1446
JS0558	JS0558	47	*Cephalotes spinosus*	4166
JS0559	JS0559	47	*Cephalotes spinosus*	1862
JS0560	JS0560	47	*Cephalotes spinosus*	4356
JS0661	JS0661	53	*Cephalotes targionii*	3213
JS0662	JS0662	53	*Cephalotes targionii*	2208
JS0663	JS0663	53	*Cephalotes targionii*	5073
JS0752	JS0752	a	*Cephalotes bohlsi*	910
JS0284	JS0284	25	*Cephalotes pallens*	1556
JS0285	JS0285	25	*Cephalotes pallens*	2458
JS0286	JS0286	25	*Cephalotes pallens*	4178
JS0585	JS0585	49	*Cephalotes pallidoides*	2388
JS0586	JS0586	49	*Cephalotes pallidoides*	4826
JS0587	JS0587	49	*Cephalotes pallidoides*	2628
JS0056	JS0056	10	*Cephalotes pellans*	921
JS0057	JS0057	10	*Cephalotes pellans*	2617
JS0058	JS0058	10	*Cephalotes pellans*	4359
JS0214	JS0214	19	*Cephalotes maculatus*	1328
JS0215	JS0215	19	*Cephalotes maculatus*	1406
JS0216	JS0216	19	*Cephalotes maculatus*	3514
JS0702.1	JS0702	lmah2	*Cephalotes maculatus*	4229
JS0703.1	JS0703	lmah2	*Cephalotes maculatus*	657
JS0705.1	JS0705	lmah2	*Cephalotes maculatus*	1251
JS0362	JS0362	21	*Cephalotes grandinosus*	795
JS0363	JS0363	21	*Cephalotes grandinosus*	5165
JS0364	JS0364	21	*Cephalotes grandinosus*	3055
JS0194	JS0194	18	*Cephalotes persimilis*	1298
JS0195	JS0195	18	*Cephalotes persimilis*	759

Table A1.2 (Continued): Number of sequences for each sample in this study.

SampleID	Individual	Colony	Species	# sequences
JS0196	JS0196	18	*Cephalotes persimilis*	1474
JS0264.1	JS0264	23	*Cephalotes persimilis*	856
JS0265	JS0265	23	*Cephalotes persimilis*	324
JS0266	JS0266	23	*Cephalotes persimilis*	550
JS0300.1	JS0300	26	*Cephalotes persimilis*	358
JS0301.1	JS0301	26	*Cephalotes persimilis*	396
JS0303.1	JS0303	26	*Cephalotes persimilis*	624
JS0755	JS0755	d	*Cephalotes crenaticeps*	2986
JS0753	JS0753	b	*Cephalotes peruviensis*	2018
JS0754	JS0754	c	*Cephalotes setulifur*	1133
JS0271	JS0271	24	*Cephalotes borgmeieri*	1273
JS0272	JS0272	24	*Cephalotes borgmeieri*	2539
JS0273	JS0273	24	*Cephalotes borgmeieri*	2152
JS0573	JS0573	48	*Cephalotes cordatus*	3251
JS0574	JS0574	48	*Cephalotes cordatus*	3185
JS0575	JS0575	48	*Cephalotes cordatus*	2842
JS0227	JS0227	20	*Cephalotes eduarduli*	2592
JS0228	JS0228	20	*Cephalotes eduarduli*	1996
JS0229	JS0229	20	*Cephalotes eduarduli*	4878
				total: 241519

Table A1.3: Alpha diversity values (mean and standard error) under UCLUST and CDHit OTU clustering methods.

			UCLUST		CDHit	
Metric	**Genus**	**Tissue**	**mean**	**SE**	**mean**	**SE**
Chao1	Azteca	gut	55.68	17.08	74.83	5.83
	Cephalotes	gut	28.02	14.08	30.26	17.64
	Cephalotes	gaster	50.17	35.64	52.78	37.24
	Procryptocerus	gaster	32.00	-	34.00	-
	Pseudomyrmex	gut	101.98	48.42	110.16	65.60
Observed Species	Azteca	gut	50.50	18.50	53.00	19.00
	Cephalotes	gut	20.18	7.05	20.16	6.93
	Cephalotes	gaster	33.67	11.16	34.17	11.04
	Procryptocerus	gaster	30.00	-	29.00	-
	Pseudomyrmex	gut	83.67	47.17	86.33	50.68
Phylogenetic Diversity	Azteca	gut	5.51	1.33	5.78	1.33
	Cephalotes	gut	2.24	0.57	2.26	0.63
	Cephalotes	gaster	3.17	0.87	3.23	0.92
	Procryptocerus	gaster	2.54	-	2.60	-
	Pseudomyrmex	gut	7.38	3.52	7.24	4.00
Shannon	Azteca	gut	2.51	1.66	2.48	1.64
	Cephalotes	gut	1.26	0.86	1.26	0.87
	Cephalotes	gaster	3.54	0.46	3.49	0.45
	Procryptocerus	gaster	2.75	-	2.79	-
	Pseudomyrmex	gut	3.21	2.12	3.27	2.16

Table A1.4: Breakdown of changes made to our dataset during the denoising process in AmpliconNoise, after Gaspar and Thomas (2013).

	Ref. step	Query step	Bases Analyzed	3' gap (bp)	Substitutions (bp)	Insertions (bp)	Deletions (bp)	Total (bp)	Error rate (%)
Filtering	0	1	73,204,027	-67	706	11,123	4,187	16,016	0.02
PyroNoise	1	2	95,168,151	92	28,256	74,159	107,755	210,170	0.22
"Accordion effect"	0	2	95,168,151	25	64,509	177,100	179,858	421,467	0.44
Truncation	2	3	88,579,026	-28	0	0	0	0	0.00
Seqnoise	3	4	89,762,805	5	46,710	46,577	36,979	130,266	0.15
Net results	0	4	89,762,805	25	103,719	189,806	189,291	479,816	0.53

Appendix I Figures

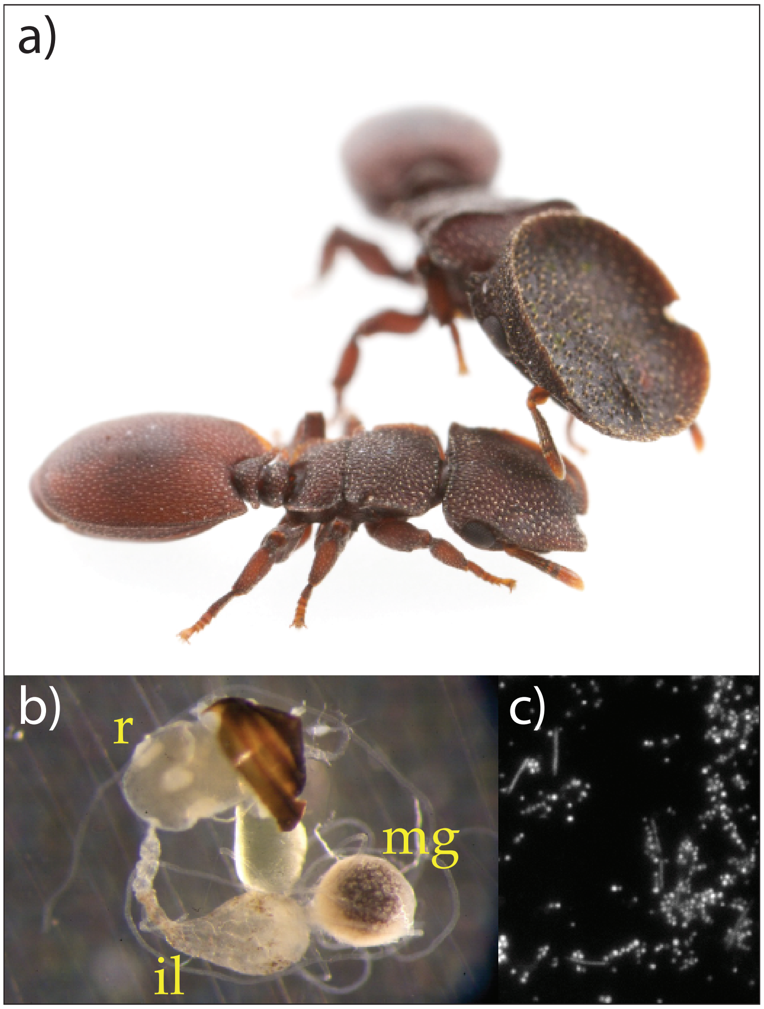


Figure A1.1: a) *Cephalotes pallidoides* worker and soldier (top). The elaborated, disc-shaped head of the soldier caste functions as a 'door' to protect the nest entrance. b) *Cephalotes* gastrointestinal tract. Clockwise from right: midgut (mg), ileum (il), rectum (r). The straw-colored organ at center is not part of the GI tract. The midgut and ileum harbor dense populations of microbes; the bulbous, enlarged ileum is unusual in ants

(de Andrade & Baroni Urbani 1999). c) Fluorescence micrograph of bacteria from *Cephalotes* midgut, illustrating morphological diversity. Rod-shaped bacteria are approximately 3μm long. All photos © Jon G Sanders.

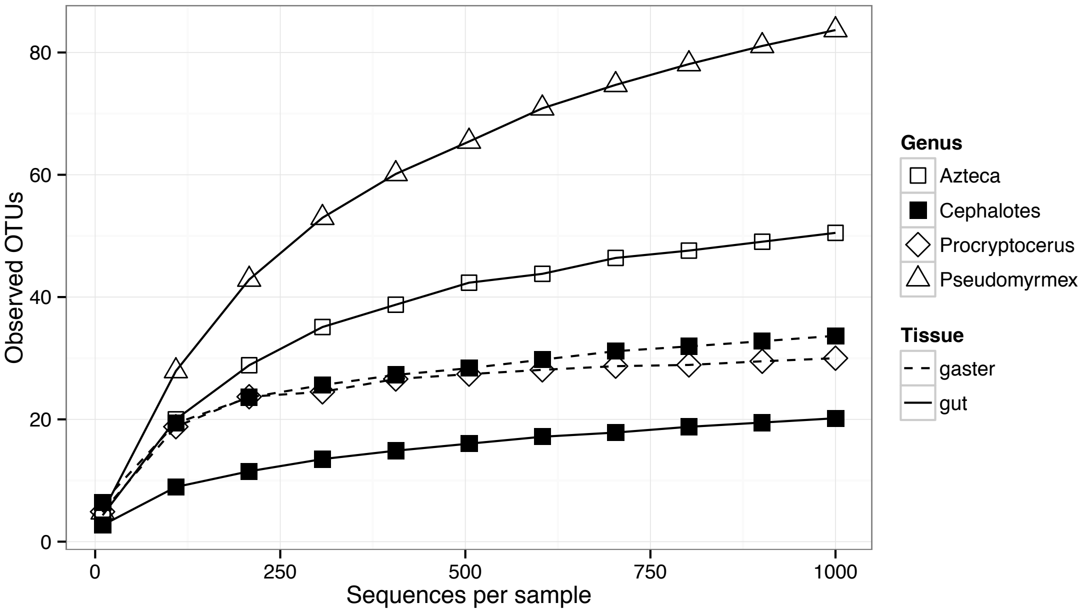


Figure A1.2: Average alpha diversity rarefaction curves for different categories of samples. *Cephalotes* samples represented by filled symbols, other genera by open symbols. Samples derived from whole ethanol-preserved gasters represented by dotted lines, those from guts by solid lines.

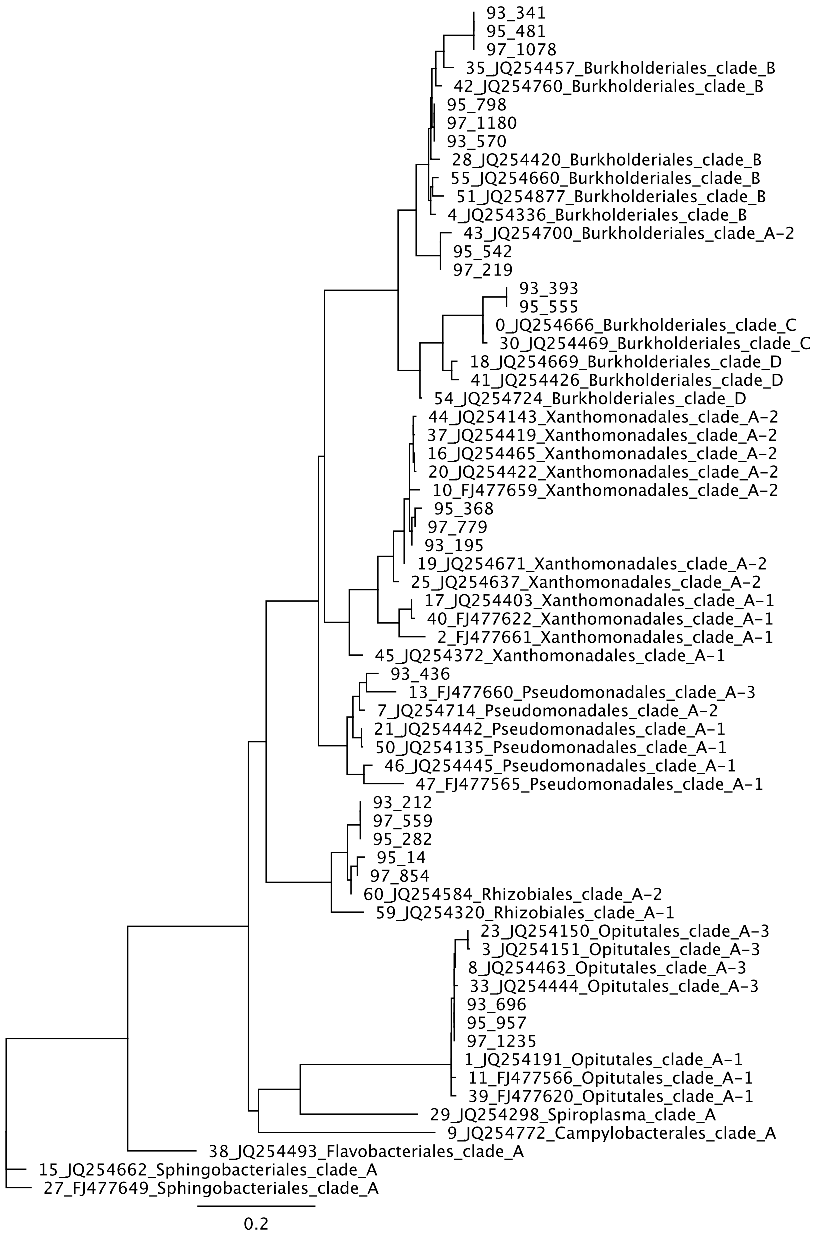

93_341
95_481
97_1078
35_JQ254457_Burkholderiales_clade_B
42_JQ254760_Burkholderiales_clade_B
95_798
97_1180
93_570
28_JQ254420_Burkholderiales_clade_B
55_JQ254660_Burkholderiales_clade_B
51_JQ254877_Burkholderiales_clade_B
4_JQ254336_Burkholderiales_clade_B
43_JQ254700_Burkholderiales_clade_A-2
95_542
97_219
93_393
95_555
0_JQ254666_Burkholderiales_clade_C
30_JQ254469_Burkholderiales_clade_C
18_JQ254669_Burkholderiales_clade_D
41_JQ254426_Burkholderiales_clade_D
54_JQ254724_Burkholderiales_clade_D
44_JQ254143_Xanthomonadales_clade_A-2
37_JQ254419_Xanthomonadales_clade_A-2
16_JQ254465_Xanthomonadales_clade_A-2
20_JQ254422_Xanthomonadales_clade_A-2
10_FJ477659_Xanthomonadales_clade_A-2
95_368
97_779
93_195
19_JQ254671_Xanthomonadales_clade_A-2
25_JQ254637_Xanthomonadales_clade_A-2
17_JQ254403_Xanthomonadales_clade_A-1
40_FJ477622_Xanthomonadales_clade_A-1
2_FJ477661_Xanthomonadales_clade_A-1
45_JQ254372_Xanthomonadales_clade_A-1
93_436
13_FJ477660_Pseudomonadales_clade_A-3
7_JQ254714_Pseudomonadales_clade_A-2
21_JQ254442_Pseudomonadales_clade_A-1
50_JQ254135_Pseudomonadales_clade_A-1
46_JQ254445_Pseudomonadales_clade_A-1
47_FJ477565_Pseudomonadales_clade_A-1
93_212
97_559
95_282
95_14
97_854
60_JQ254584_Rhizobiales_clade_A-2
59_JQ254320_Rhizobiales_clade_A-1
23_JQ254150_Opitutales_clade_A-3
3_JQ254151_Opitutales_clade_A-3
8_JQ254463_Opitutales_clade_A-3
33_JQ254444_Opitutales_clade_A-3
93_696
95_957
97_1235
1_JQ254191_Opitutales_clade_A-1
11_FJ477566_Opitutales_clade_A-1
39_FJ477620_Opitutales_clade_A-1
29_JQ254298_Spiroplasma_clade_A
9_JQ254772_Campylobacterales_clade_A
38_JQ254493_Flavobacteriales_clade_A
15_JQ254662_Sphingobacteriales_clade_A
27_FJ477649_Sphingobacteriales_clade_A
0.2

Figure A1.3: Maximum Likelihood tree of "core" *Cephalotes* OTUs (as defined by presence in ≥50% of *Cephalotes* samples) and representative "*Cephalotes*-specific" OTUs (one representative sequence per 97% OTU as published in Anderson *et al.* [2012]). ML tree computed with alignments generated in PYNAST against the same Greengenes database used for the primary analysis, and filtered of hypervariable sites using the associated lanemask. Filtered sequences were then imported into Geneious v6.1.6 (Biomatters Ltd) and the maximum likelihood tree computed using PHYLIP and a GTR model of sequence evolution.

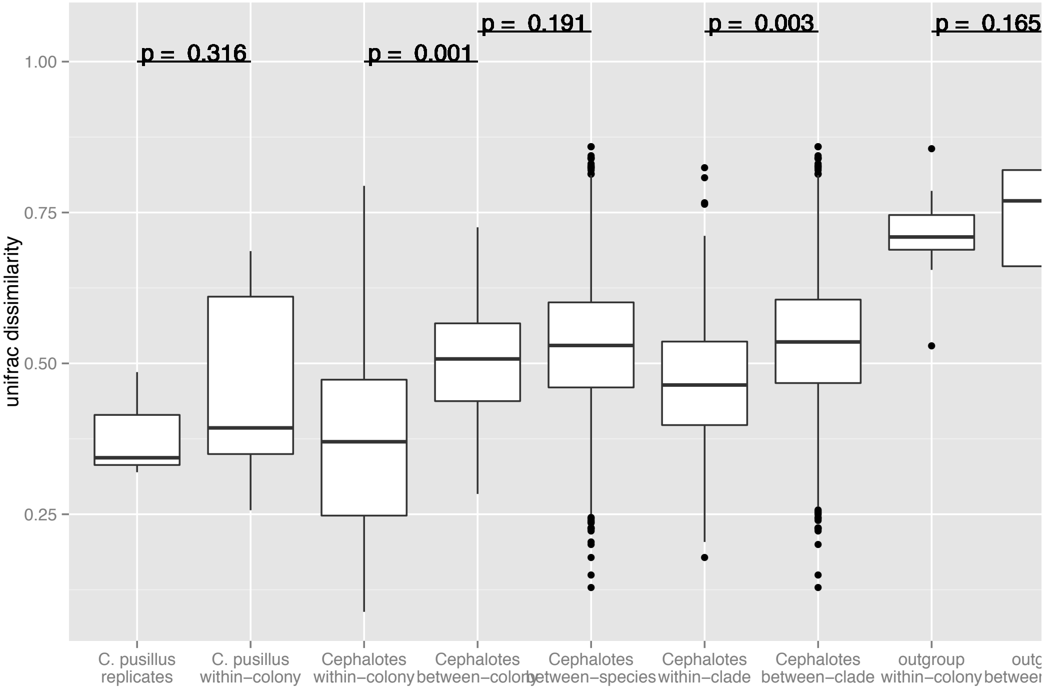


Figure A1.4: Dissimilarity boxplots, comparing community Unifrac distances within and among groups of ants. Significance values for between-group comparisons calculated using Monte Carlo permutation tests (see Appendix I for details on permutation structure).

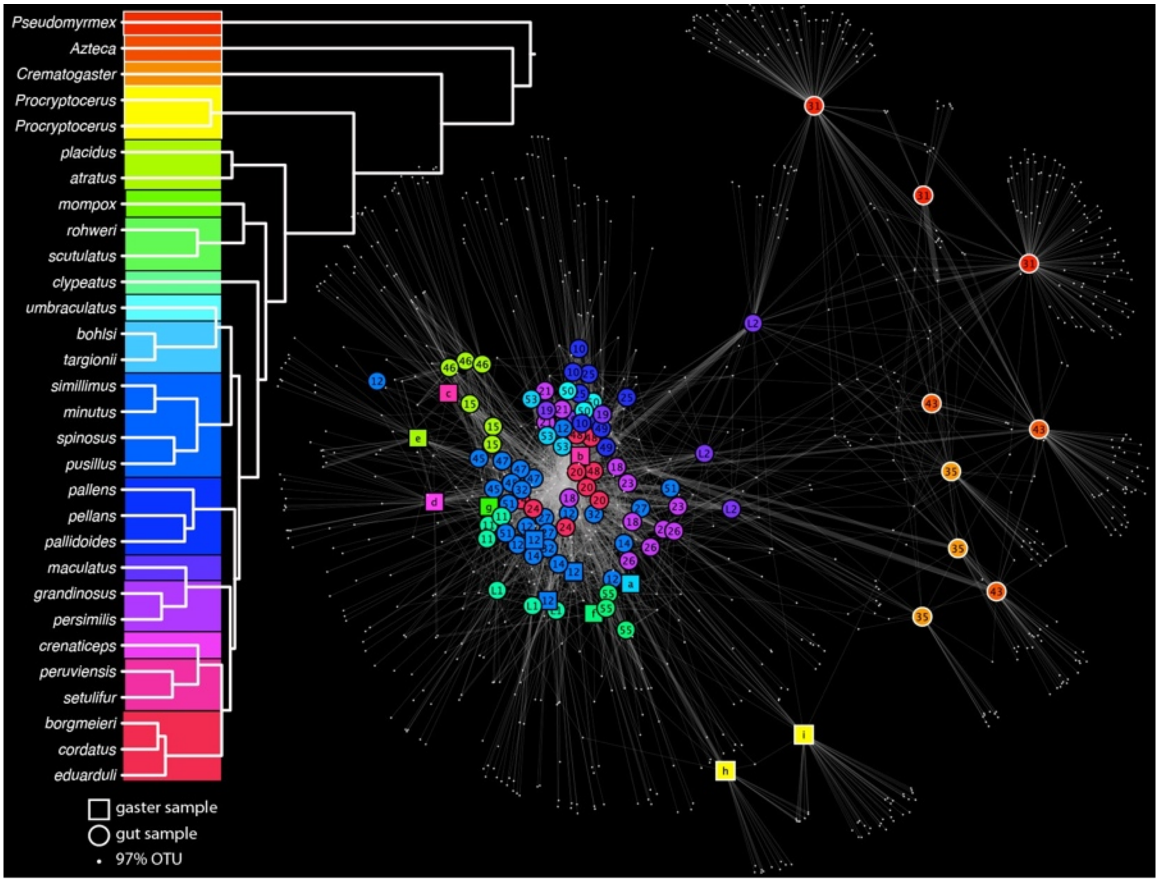

Pseudomyrmex
Azteca
Crematogaster
Procryptocerus
Procryptocerus
placidus
atratus
mompox
rohweri
scutulatus
clypeatus
umbraculatus
bohlsi
targionii
simillimus
minutus
spinosus
pusillus
pallens
pellans
pallidoides
maculatus
grandinosus
persimilis
crenaticeps
peruviensis
setulifur
borgmeieri
cordatus
eduarduli
□ gaster sample
○ gut sample
• 97% OTU

Figure A1.5: OTU network showing relationships among ant gut microbiota. Large, colored nodes represent individual ant samples, while small grey nodes represent individual 97% OTUs. Edges connect OTUs with each host sample in which they occur. Nodes are placed according to a weighted, spring-embedded algorithm, causing host nodes that share more OTUs to appear close together. Host nodes are labeled by colony (see Table 1) and colored by clades (sensu De Andrade and Baroni-Urbani [1999]) in rainbow order according to phylogeny (inset), with non-*Cephalotes* host nodes outlined in white for emphasis

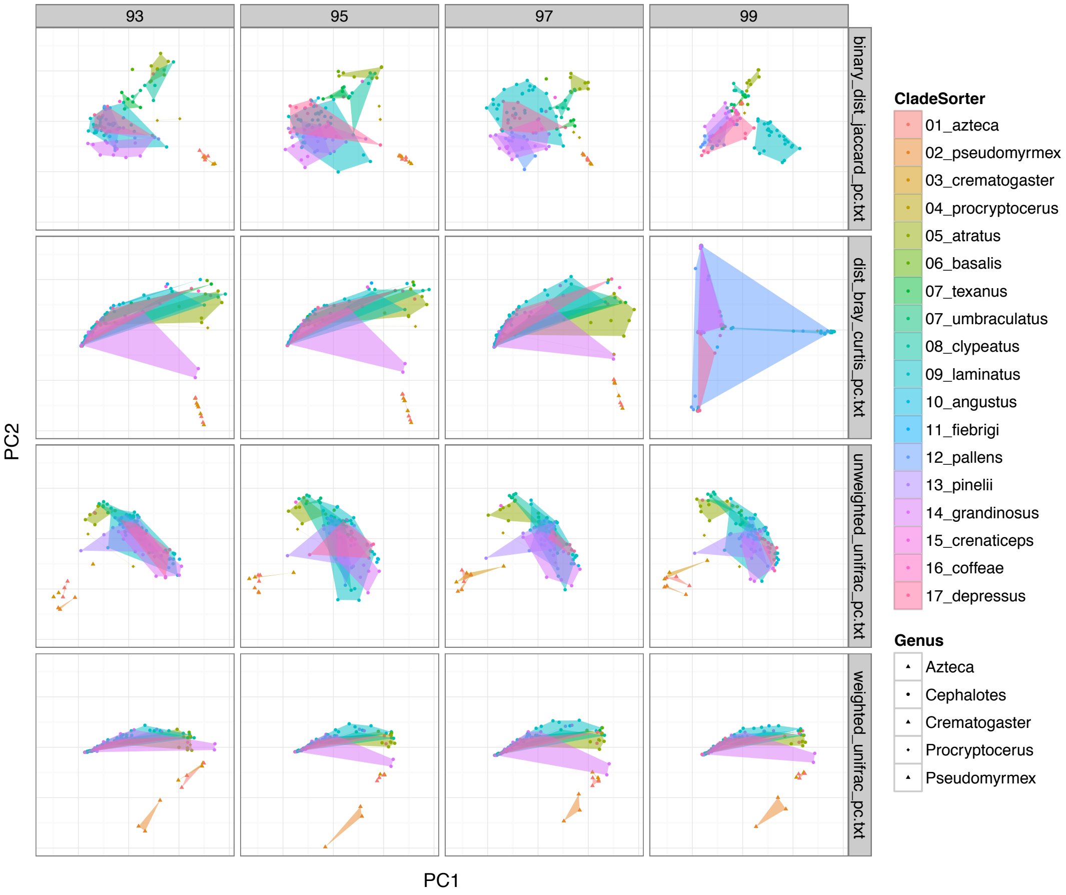


Figure A1.6: PCoA plots showing relationships among samples, for four different beta diversity measures (Jaccard, Bray-Curtis, Unweighted UniFrac, and Weighted UniFrac) and four OTU clustering widths (93, 95, 97, and 99% identity). Samples are indicated with different shapes for each genus (solid triangles for non-cephalotine genera *Azteca*, *Crematogaster*, and *Pseudomyrmex*, diamonds for *Procryptocerus*, and circles for *Cephalotes*) and colored by host clade. Additionally, a minimum convex polygon is plotted to highlight groupings for each host clade, and colored accordingly.

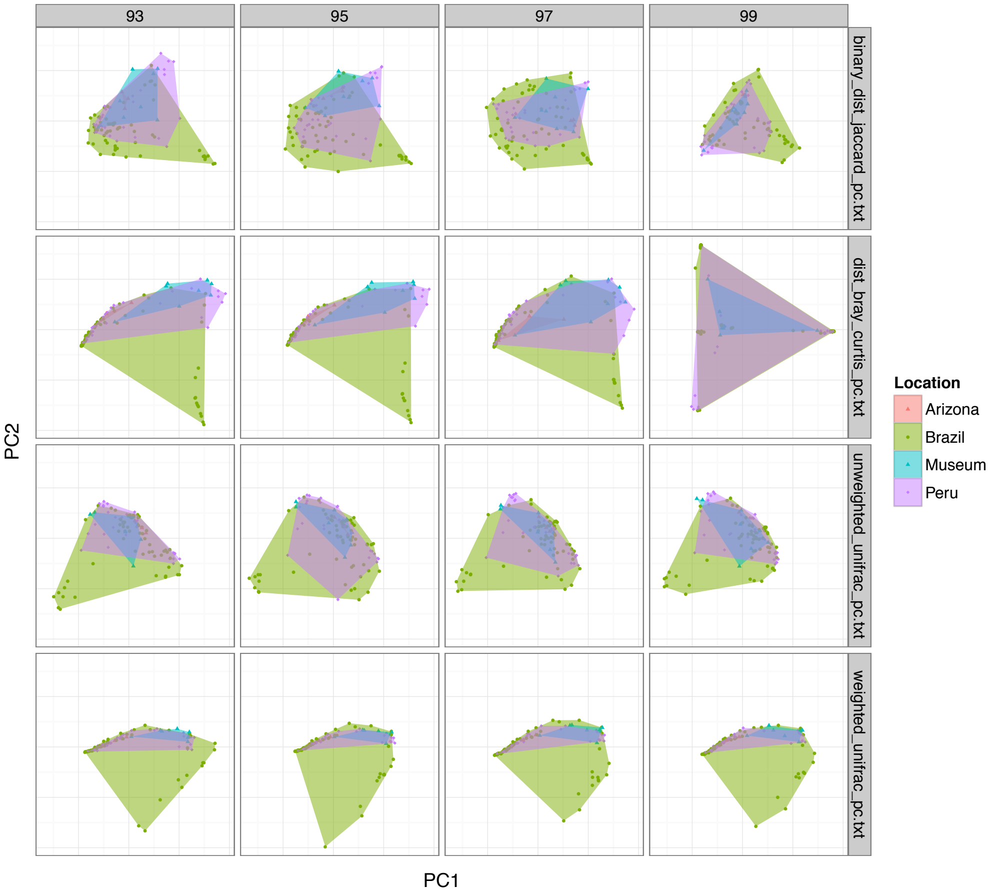


Figure A1.7: As in Figure A1.6, but with points and polygons colored according to location of origin.

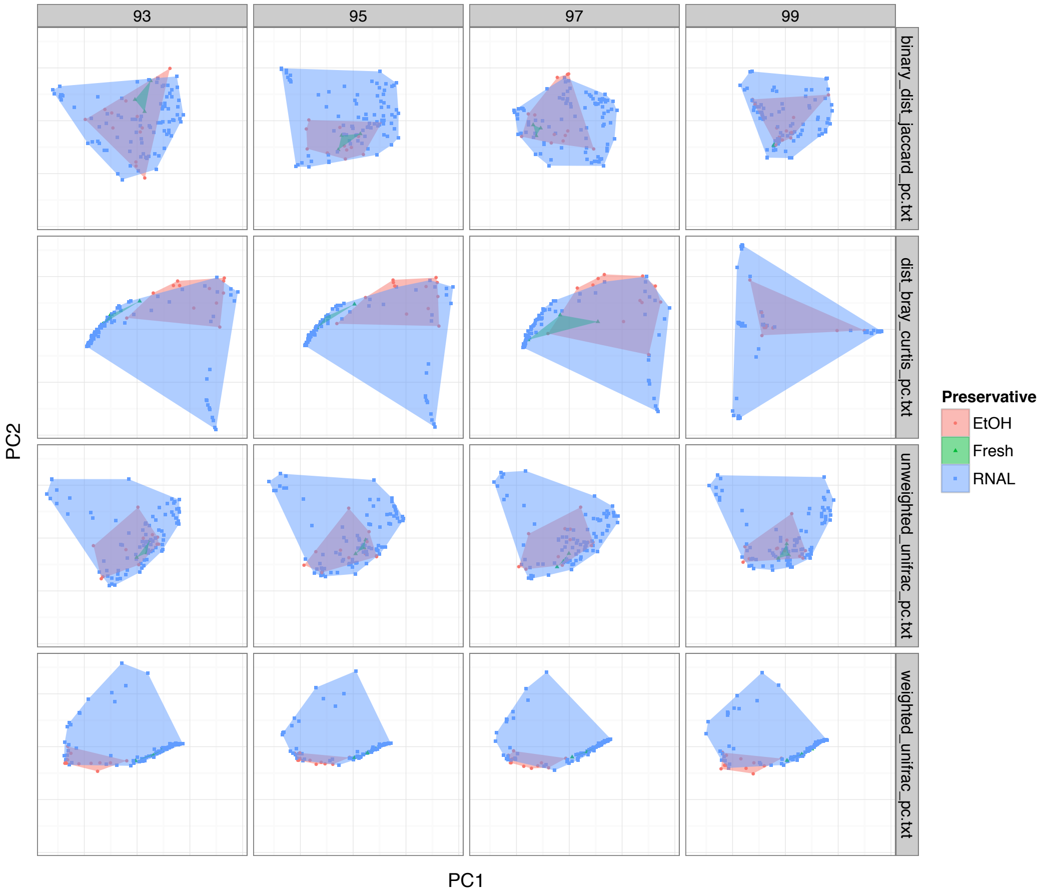


Figure A1.8: As in Figure A1.6, but with points and polygons colored according to preservative.

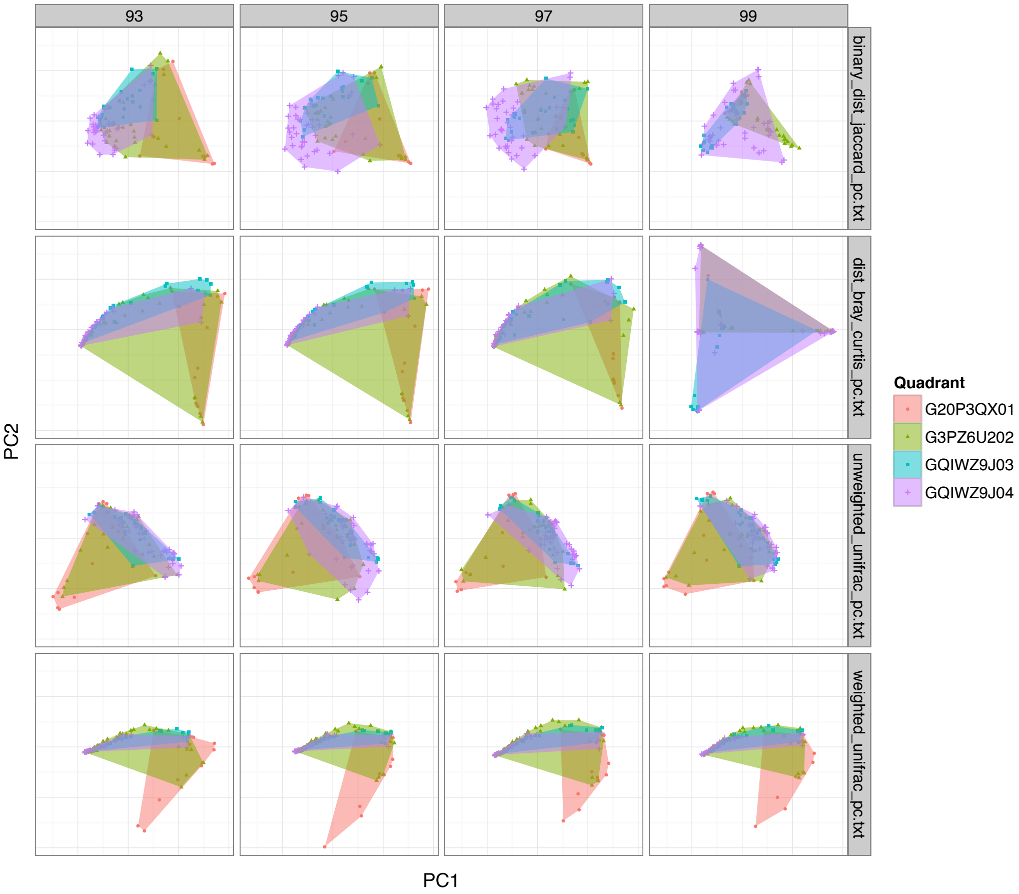


Figure A1.9: As in Figure A1.6, but with points and polygons colored according to sequencing quadrant.

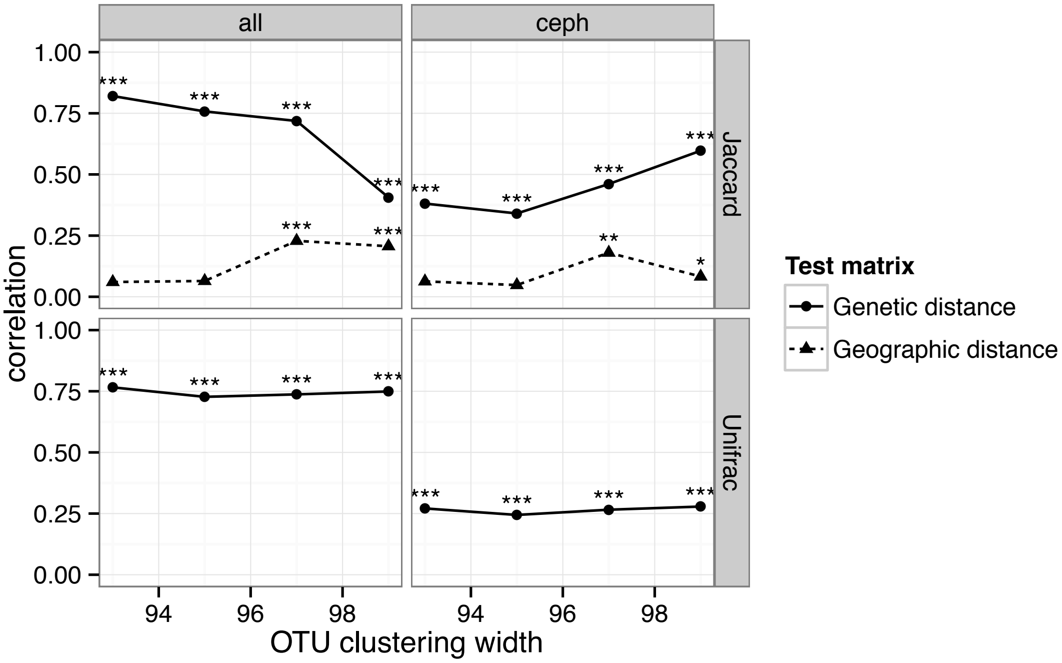


Figure A1.10: Results from partial Mantel tests show the influence of phylogeny in explaining variation among ant microbiomes. Correlation coefficients between UniFrac distances and Jaccard dissimilarity matrices and genetic and geographic distances, respectively, are plotted at four different OTU clustering thresholds. Results for Jaccard dissimilarities are repeated from Fig. 5 for comparison to UniFrac results. Results plotted for *Cephalotes*-only ('ceph') and for *Cephalotes* plus outgroups ('all'). P-values from partial Mantel tests indicated by asterisks (***, $p = 0.001$; **, $0.001 < p < 0.01$).

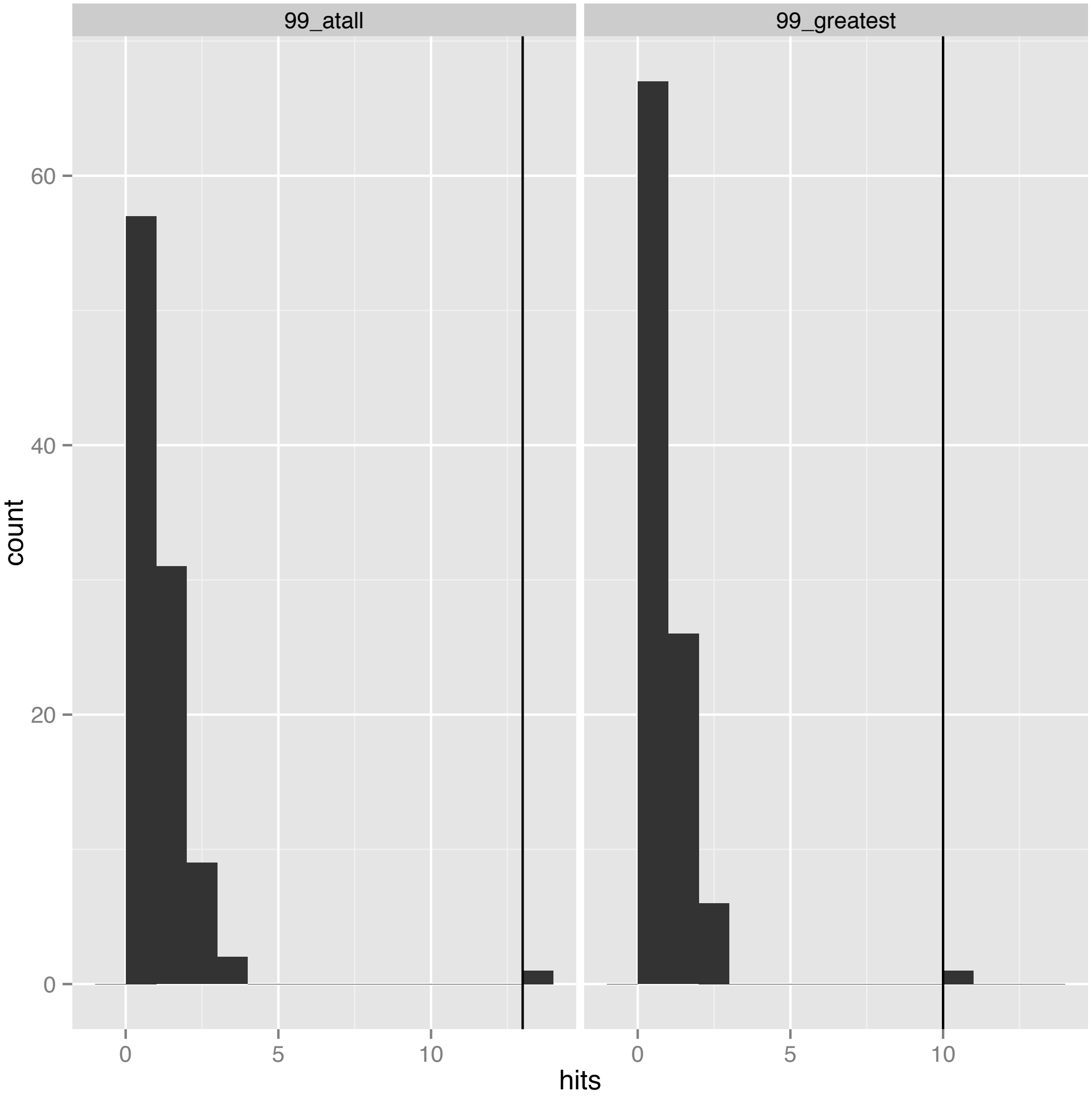


Figure A1.11: Histograms showing the results of beta diversity sensitivity analyses using host trees with randomly permuted tips for comparison. Histograms show frequency distributions of the number of clades recovered at all under 99% Jaccard similarity (left panel) and the number of clades recovered for which the 99% Jaccard value was higher than that for any wider clustering threshold with the same diversity measure (right panel). The number of nodes recovered from the unpermuted tree is included in each histogram, denoted by a vertical line.

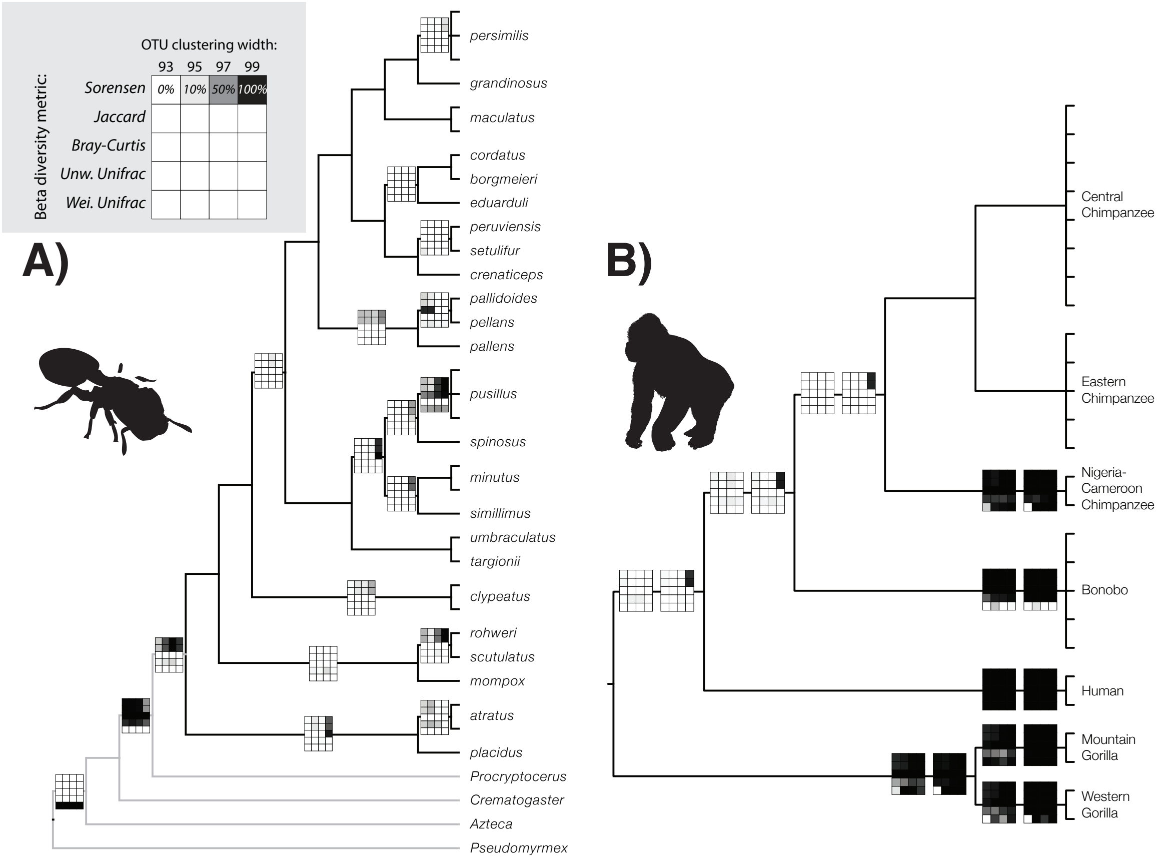

OTU clustering width:
93
95
97
99
Beta diversity metric:
Sorensen
0%
10%
50%
100%
Jaccard
Bray-Curtis
Unw. Unifrac
Wei. Unifrac
A)
persimilis
grandinosus
maculatus
cordatus
borgmeieri
eduarduli
peruviensis
setulifur
crenaticeps
pallidoides
pellans
pallens
pusillus
spinosus
minutus
simillimus
umbraculatus
targionii
clypeatus
rohweri
scutulatus
mompox
atratus
placidus
Procryptocerus
Crematogaster
Azteca
Pseudomyrmex
B)
Central Chimpanzee
Eastern Chimpanzee
Nigeria-Cameroon Chimpanzee
Bonobo
Human
Mountain Gorilla
Western Gorilla

Figure A1.12: Version of Figure 1.5 with additional beta diversity indices noted in legend

APPENDIX 2

Chapter 2 Supplemental

Odontocetes

In contrast to baleen whales, the toothed whale samples in our dataset were highly variable in taxonomic composition, often dominated by one or two bacterial lineages typically found at lower abundance in other mammalian guts (Fig 2.1a). Some of these dominant OTUs were perfect matches to sequences previously recovered from marine environmental samples, including the Gammaproteobacteria *Pseudoalteromanas* and *Photobacterium*, suggesting the potential for an environmental origin. Some of these sequences also matched a library of bacterial 16S sequences associated with marine copepods, further suggesting the potential impact of environmental bacteria (Fig A2.3). However, three other dominant OTUs were 99-100% matches to sequences previously recovered from other marine mammals, including *Mycoplasma* (Tenericutes), *Mycobacterium* (Actinobacteria), and *Cetobacterium* (Fusobacteria). In aggregate, the taxonomic composition we observed in odontocetes showed a similar trend to what has previously been reported for piscivorous pinnipeds (Nelson *et al.* 2012; 2013), with an enrichment in Proteobacteria and Fusobacteria relative to terrestrial mammals.

The three odontocete metagenomes also showed notable differences to the overall pattern of similarity with 16S data, with these samples having dramatically different profiles of predicted functional genes. Taxonomic profiling suggests an extremely high degree of host genomic contamination for these samples: between 20 and 90% of reads matched most closely to eukaryotic sequences in the KEGG database (Fig A2.9). Manual inspection of these reads reveals that most match closely to the low-coverage genomic sequence available for the dolphin *Trusiops truncatus* (Ensembl accession *turTru1*) or to *Bos taurus*, suggesting that the sequenced DNA was derived from the host animal rather than undigested food or laboratory contamination. The source of this host DNA is unknown, but may arise from sloughing intestinal epithelia in a midgut that is much longer than in most other mammals (Williams *et al.* 2001). Together with the high taxonomic variability observed in 16S data from these samples, the high proportion of host sequence suggests that these samples may not be directly comparable to other samples; we have consequently excluded them from further analysis.

Difference among whales

Comparison of whale species within our sample set also showed moderate differences in microbial composition, potentially reflecting differences between diets. Our Illumina 16S data consisted of seven right whales, which feed primarily upon crustaceans such as calanoid copepods, and five rorquals, which consume both vertebrate and invertebrate prey. Compared with right whales, rorquals appeared to retain more of the bacterial taxa that were also more abundant in terrestrial mammals, such as Bacilli, *Blautia*, *Coprococcus*, and *Coprobacillus,* than did right right whales. Diets consisting entirely of invertebrates would be likely to contain higher proportions of fermentable animal polysaccharides (including chitin). Consistent with this, the clade 5 Verrucomicrobia, which were more enriched in baleen whales and terrestrial herbivores than in other mammals, were also significantly more abundant in right whales than in rorquals. These overall differences in community composition were especially apparent in an initial 454 dataset of the V1-V3 regions of 16S, with baleen whales grouping separately from both terrestrial samples and from Antarctic seals (Fig A2.10).

Table A2.1: Detailed collection information for samples in this study.

Host ID	Host Species	Collection Date	Collection Location	454 16S (V1-V3)	Illm 16S (V4)	Meta
JS1	Sei Whale (*Balaenoptera borealis*)	9/18/11	Grand Manan, NB, Canada	Yes	Yes	
JS2	Atlantic Whitesided Dolphin (*Lagenorhynchus acut*	8/23/11	Grand Manan, NB, Canada	Yes	Yes	
JS3	Coyote (*Canis latrans*)	11/15/12	Concord, MA	Yes	Yes	Yes
JS4	Mouse (*Mus musculus*)	12/14/12	Harvard University, MA	Yes	Yes	Yes
JS5	Horse (*Equus ferus caballus*)	12/9/12	Concord, MA	Yes	Yes	Yes
JS6	White-tailed Deer (*Odocoileus virginianus*)	11/25/12	Concord, MA	Yes	Yes	Yes
JS7	Fisher (*Martes pennanti*)	11/22/12	Stow, MA	Yes	Yes	Yes
JS8	Rabbit (*Oryctolagus cuniculus*)	12/13/12	Cambridge, MA	Yes	Yes	Yes
JS9	Humback Whale (*Megaptera novaeangliae*)	10/12/08	Seymour Canal, AK		Yes	
JS10	Humback Whale (*Megaptera novaeangliae*)	6/22/11	Sitka Sound, AK		Yes	Yes
JS11	Humback Whale (*Megaptera novaeangliae*)	11/15/09	Seymour Canal, AK		Yes	
JS12	Humback Whale (*Megaptera novaeangliae*)	11/15/09	Seymour Canal, AK		Yes	
JS13	Bottlenose Dolphin (*Tursiops truncatus)*	4/22/11	Long Marine Laboratory, CA		Yes	Yes
JS14	Bottlenose Dolphin (*Tursiops truncatus)*	4/22/11	Long Marine Laboratory, CA		Yes	
JS16	Beluga (*Delphinapterus leucas*)	5/28/11	Mystic Aquarium, CT		Yes	
JS17	Beluga (*Delphinapterus leucas*)	5/28/11	Mystic Aquarium, CT		Yes	Yes
JS18	Hippopotamus (*Hippopotamus amphibius*)	8/22/12	Kenya		Yes	
JS19	Hippopotamus (*Hippopotamus amphibius*)	8/22/12	Kenya		Yes	Yes
JS20	Hippopotamus (*Hippopotamus amphibius*)	8/21/12	Kenya		Yes	
JS21	Hippopotamus (*Hippopotamus amphibius*)	8/21/12	Kenya		Yes	
F2	Right Whale (*Eubalaena glacialis*)	8/13/11	Grand Manan, NB, Canada		Yes	
F5	Right Whale (*Eubalaena glacialis*)	8/13/11	Grand Manan, NB, Canada	Yes	Yes	
F8	Right Whale (*Eubalaena glacialis*)	8/23/11	Grand Manan, NB, Canada	Yes	Yes	
F9	Right Whale (*Eubalaena glacialis*)	8/24/11	Grand Manan, NB, Canada		Yes	Yes
F11	Right Whale (*Eubalaena glacialis*)	8/24/11	Grand Manan, NB, Canada		Yes	Yes
F12	Right Whale (*Eubalaena glacialis*)	8/24/11	Grand Manan, NB, Canada	Yes	Yes	Yes
F16	Right Whale (*Eubalaena glacialis*)	9/2/11	Grand Manan, NB, Canada	Yes	Yes	Yes

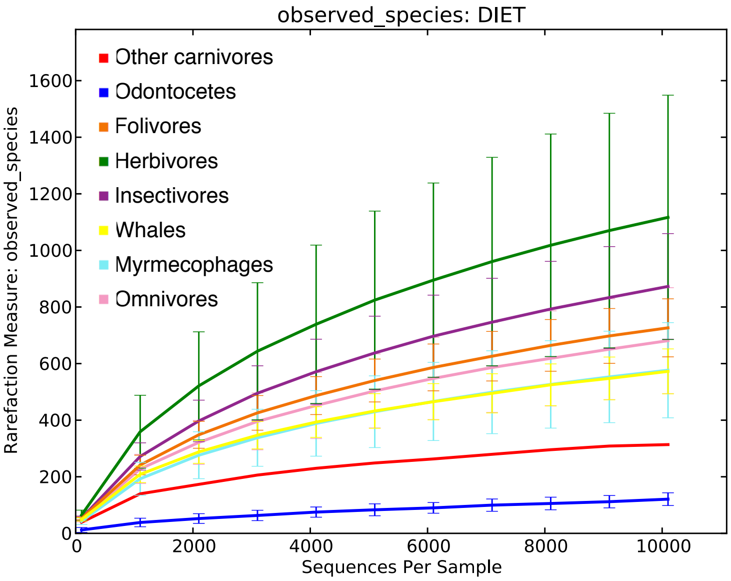


Fig. A2.2: 16S rRNA rarefaction curves indicating mean alpha diversity (observed 97% OTUs) for different mammalian dietary categories, error bars indicating standard deviations. Odontocetes had the lowest observed diversity, followed by non-insectivorous carnivores, whales, and myrmecophages.

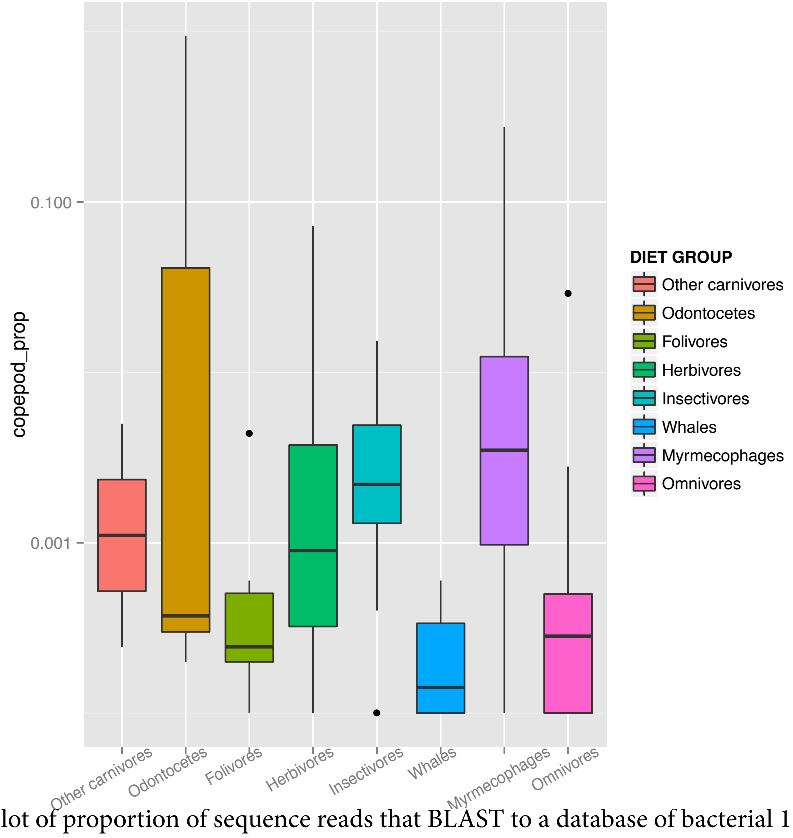


Fig. A2.3: Boxplot of proportion of sequence reads that BLAST to a database of bacterial 16S sequences associated with North Atlantic calanoid copepods, the preferred food of the right whales in this study. Y axis is log transformed.

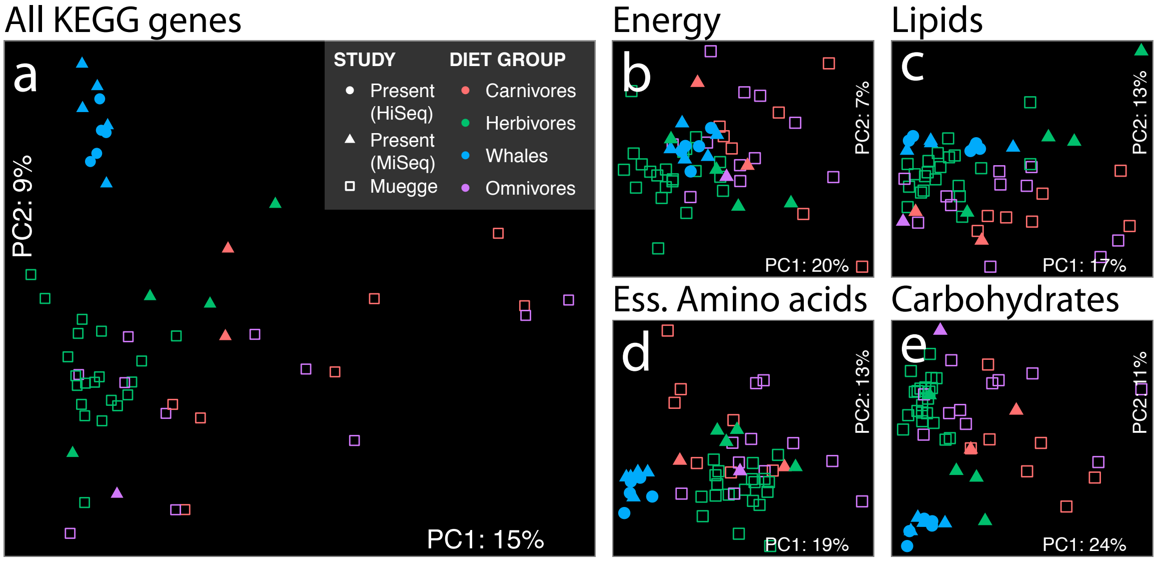


Fig. A2.4: Unconstrained RDA ordinations of KEGG gene abundances in microbiomes of whales, other carnivores, herbivores, and omnivores; identical to Figure 2, except calculated using individual abundances rather than pathway abundances. Note that in (d), whales group separately from terrestrial mammals, whereas when considering whole pathway abundances, they group together with the terrestrial carnivores.

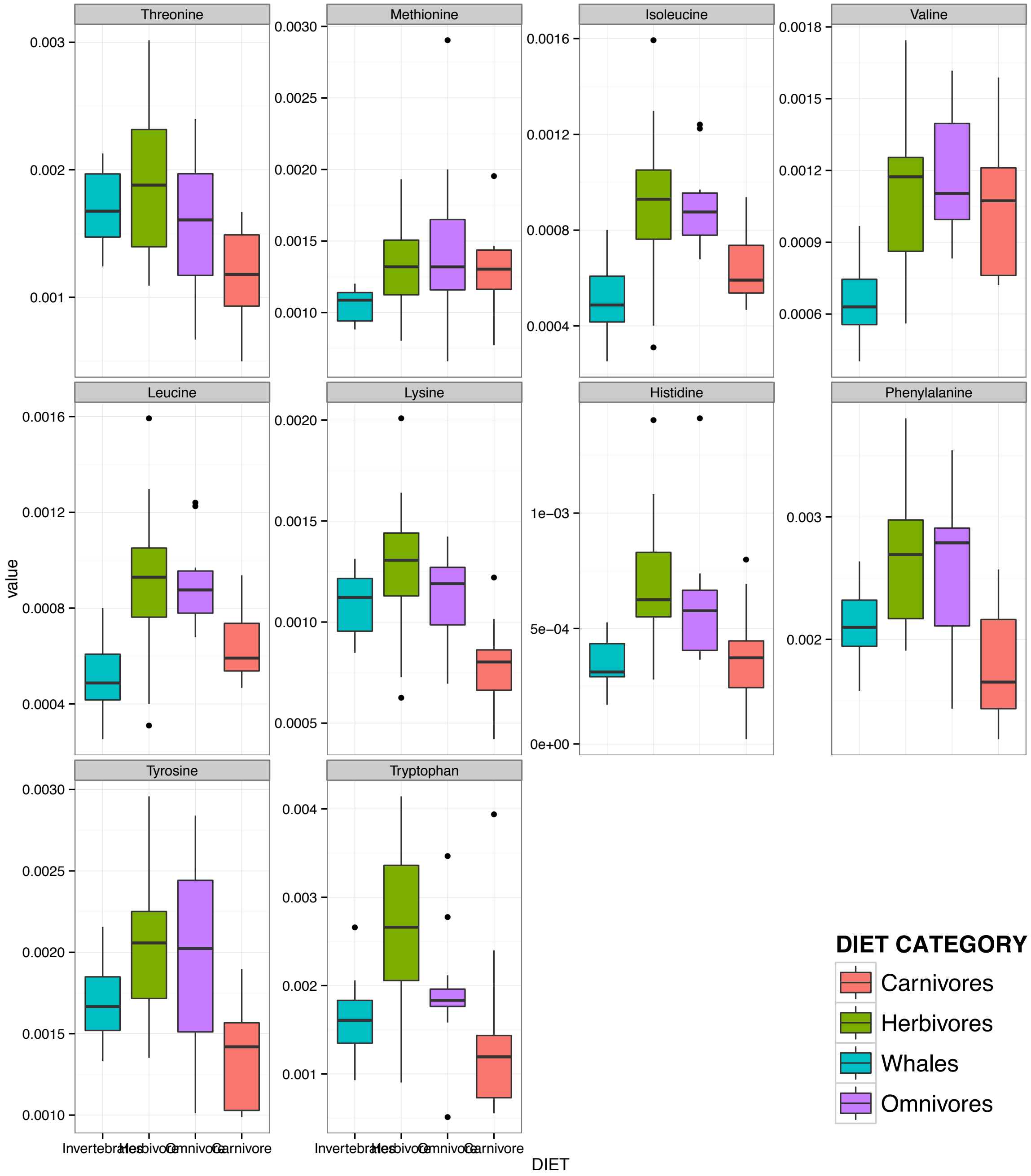


Fig. A2.5: Normalized abundances for genes catalyzing the final step of biosynthesis for the nine essential amino acids. According to Mann-Whitney post-hoc pairwise tests, whale microbiomes are significantly depleted relative to herbivore microbiomes for all but Threonine, Lysine, and Tyrosine; they are significantly enriched relative to terrestrial carnivores only for Threonine and Lysine.

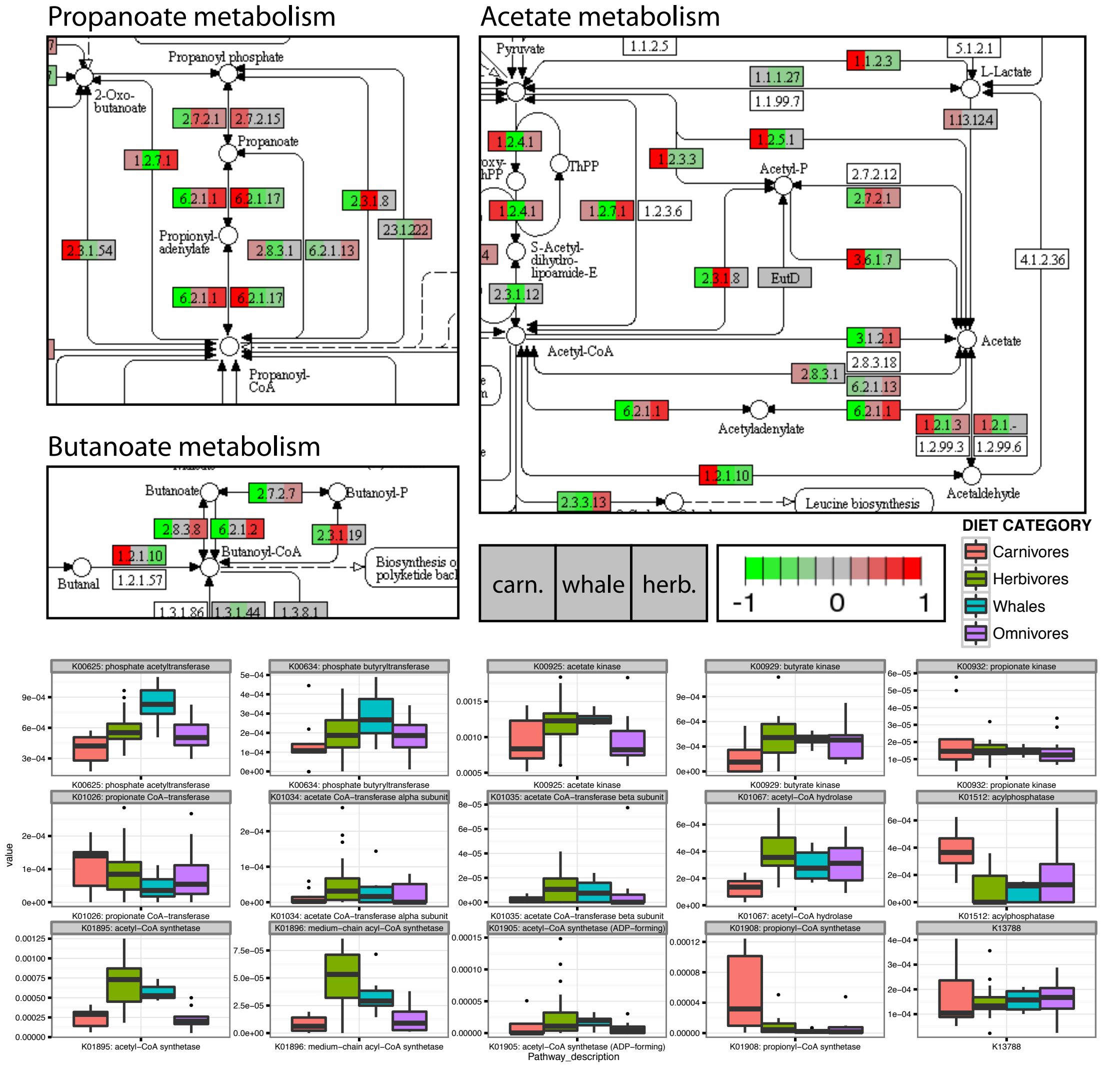


Fig. A2.6: KEGG pathway vignettes and normalized gene abundances indicating the relative enrichment of genes involved in the processing of the Short Chain Fatty Acids (SCFAs) acetate, propanoate, and butanoate. Genes performing each catalytic function (represented by E.C. numbers) are indicated for carnivores, baleen whales, and herbivores are represented for each E.C. number in that order. Gene abundances are normalized, mean-centered values, and averaged by dietary category; genes normalized including omnivore samples, but only carnivores, whales, and herbivores are displayed. Red boxes indicate that the average normalized abundance for that dietary category is greater than the global average for that gene; green, that it is less than average.

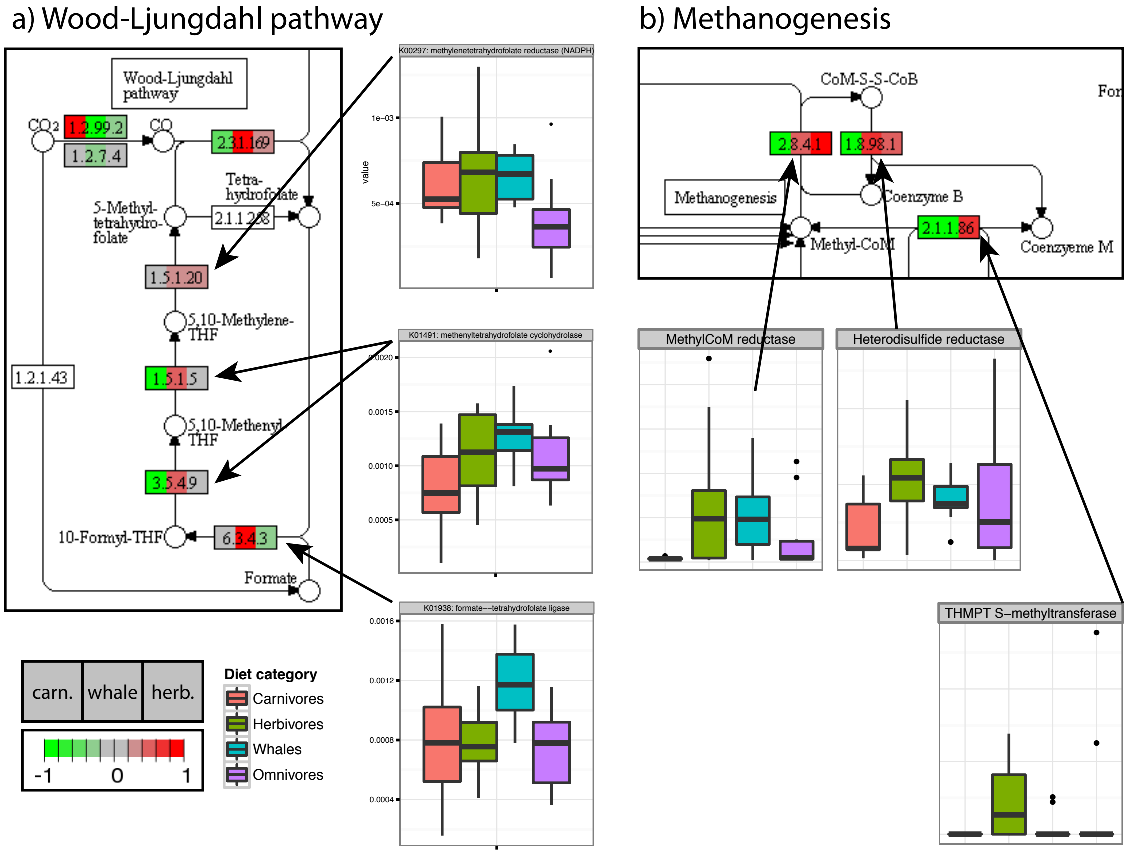


Fig. A2.7: a) Relative abundance of key genes in the Wood-Ljungdahl pathway. KEGG pathway vignette is colored as in Fig. ED6. b) Relative abundance of key genes in methanogenesis. Box plots include the summed relative abundance of all annotated subunits for each enzyme.

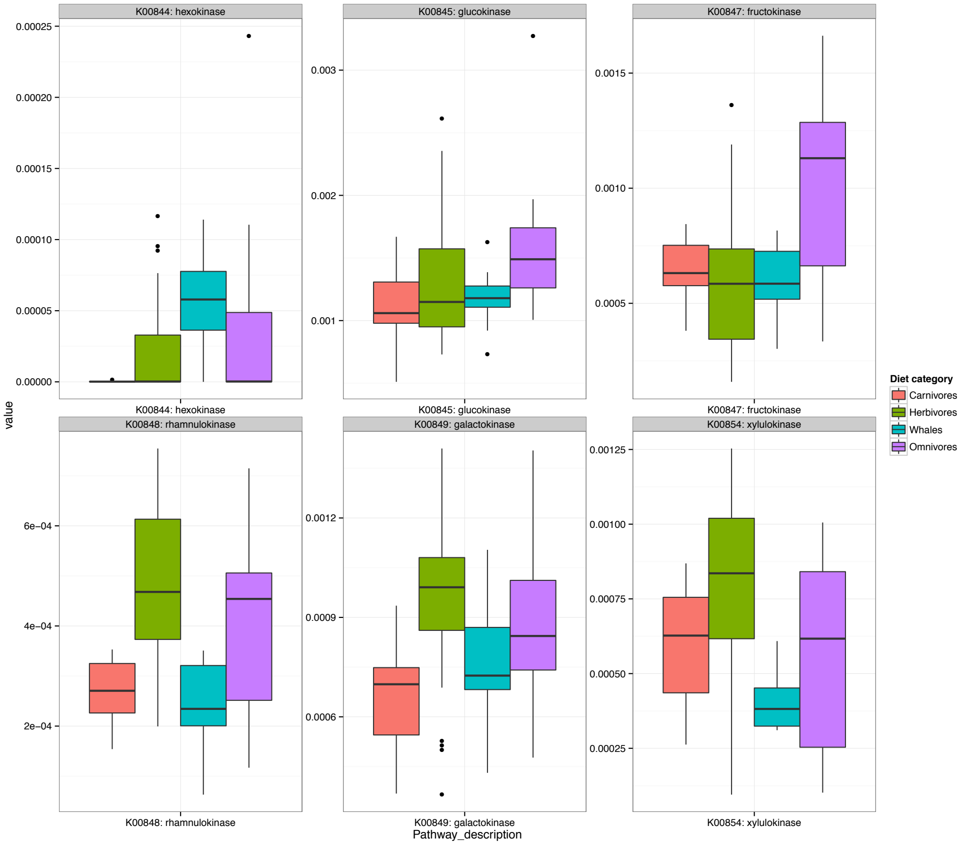


Fig. A2.8: Relative abundances of monosaccharide kinases in mammal microbiomes. Like terrestrial carnivores, whales were relatively depleted in kinases for sugars commonly found in hemicellulose.

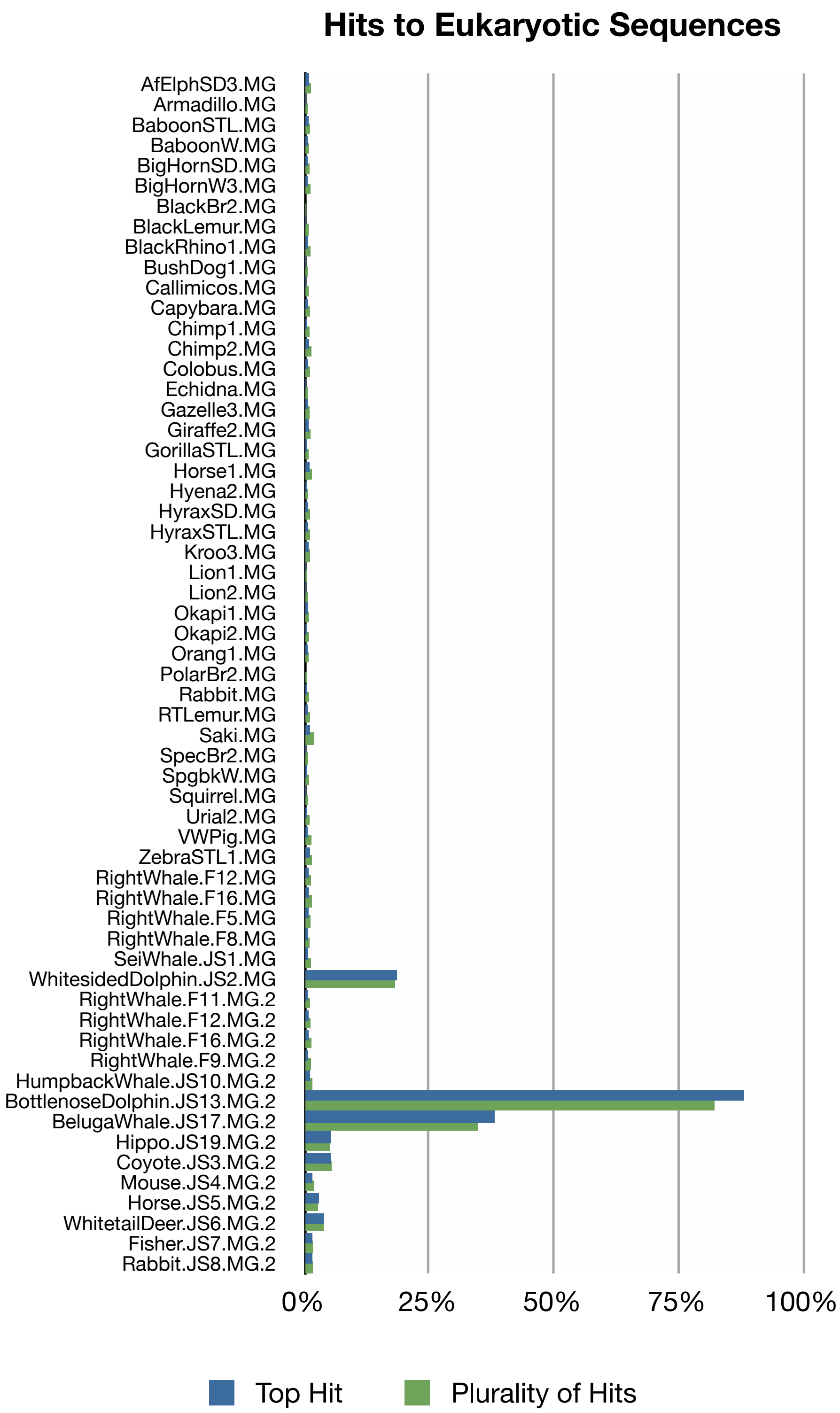


Fig. A2.9: Likely host contamination as a proportion of overall reads for each dataset. Blue bar in each pair indicates proportion of reads that matched to a eukaryotic sequences as top hit; green bar indicates proportion of reads for which a plurality of top hits were from eukaryote.

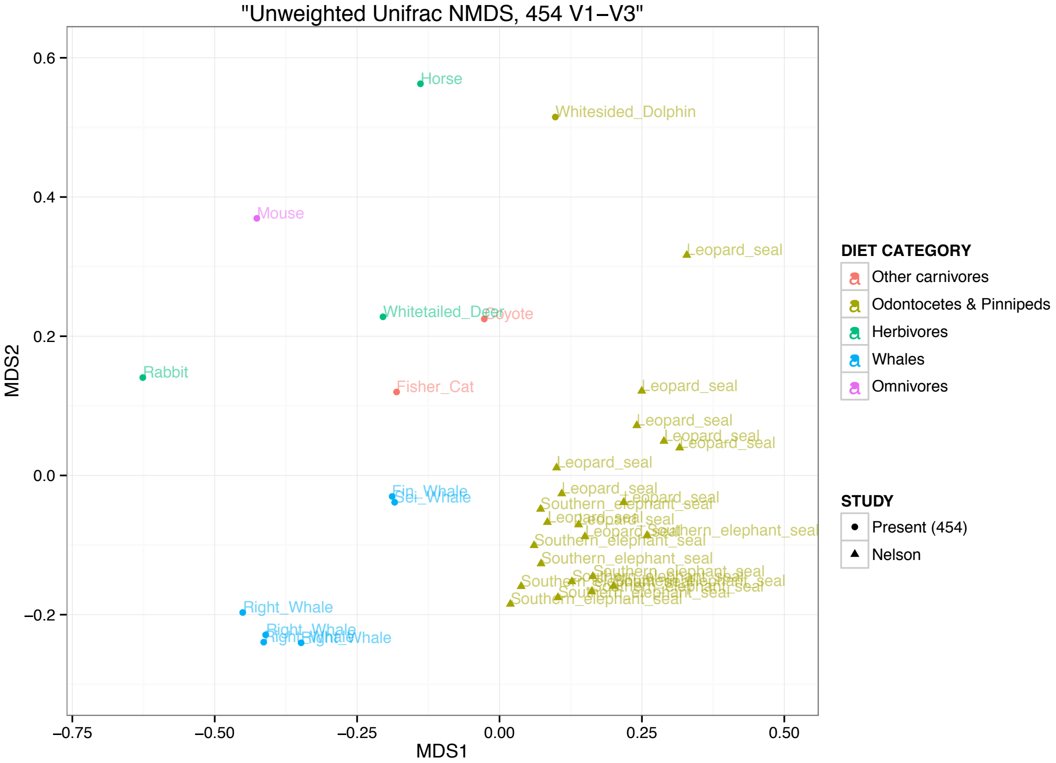


Fig. A2.10: NMDS ordination of unweighted UniFrac distances for pinniped samples from Ref. 38 and the present study. Distances based on 97% OTUs clustered from 454 sequencing of the V1-V3 region of the bacterial 16S rRNA gene.

APPENDIX 3

Chapter 3 Supplemental

Table A3.1: collections data

Colony Accession	qPCR N	∂15N	Visual Score	Habitat	Subfamily	Genus
JS-C-050	3	NA	NA	arboreal	MYRMICINAE	Cephalotes
JS-C-051	3	NA	NA	arboreal	MYRMICINAE	Cephalotes
JS-C-052	3	NA	NA	arboreal	MYRMICINAE	Cephalotes
JS-C-079	3	7.48	0	ground	MYRMICINAE	Megalomyrmex
JS-C-080	3	11.31	0	ground	PONERINAE	Neoponera
JS-C-081	3	6.19	0	arboreal	DOLICHODERINAE	Dolichoderus
JS-C-082	3	9.32	1	ground	PONERINAE	Odontomachus
JS-C-083	3	5.45	2	arboreal	FORMICINAE	Camponotus
JS-C-084	2	9.09	0	arboreal	DOLICHODERINAE	Azteca
JS-C-085	3	8.03	3	arboreal	DOLICHODERINAE	Dolichoderus
JS-C-086	3	4.05	2	arboreal	FORMICINAE	Camponotus
JS-C-087	3	9.09	3	arboreal	MYRMICINAE	Crematogaster
JS-C-088	3	11.55	0	ground	MYRMICINAE	Pheidole
JS-C-089	3	6.57	3	arboreal	DOLICHODERINAE	Dolichoderus
JS-C-090	3	8.84	0	ground	ECTATOMMINAE	Ectatomma
JS-C-091	3	9.78	0	arboreal	MYRMICINAE	Crematogaster
JS-C-092	3	6.59	3	arboreal	DOLICHODERINAE	Azteca
JS-C-093	3	6.89	0	ground	ECITONINAE	Eciton
JS-C-094	3	NA	4	arboreal	MYRMICINAE	Cephalotes
JS-C-095	3	NA	4	arboreal	DOLICHODERINAE	Dolichoderus
JS-C-096	3	3.79	1	arboreal	PSEUDOMYRMICINAE	Pseudomyrmex
JS-C-097	3	7.00	1	arboreal	PSEUDOMYRMICINAE	Pseudomyrmex
JS-C-098	3	12.28	3	ground	ECITONINAE	Neivamyrmex
JS-C-099	3	4.20	2	arboreal	DOLICHODERINAE	Dolichoderus
JS-C-100	3	4.89	0	arboreal	PSEUDOMYRMICINAE	Pseudomyrmex
JS-C-101	3	NA	3	arboreal	MYRMICINAE	Procryptocerus
JS-C-102	3	7.19	0	arboreal	MYRMICINAE	Cephalotes
JS-C-105	3	7.09	0	arboreal	FORMICINAE	Camponotus
JS-C-106	3	10.02	0	arboreal	MYRMICINAE	Crematogaster
JS-C-108	3	3.51	2	arboreal	FORMICINAE	Myrmelachista
JS-C-109	3	9.46	0	ground	MYRMICINAE	Pheidole
JS-C-110	3	10.51	0	ground	MYRMICINAE	Pheidole
JS-C-111	3	12.95	0	ground	PONERINAE	Pseudoponera
JS-C-113	2	11.50	2	ground	ECITONINAE	Eciton
JS-C-114	3	15.25	0	ground	PONERINAE	Neoponera
JS-C-116	3	10.55	0	ground	MYRMICINAE	Pheidole
JS-C-117	3	11.12	0	ground	MYRMICINAE	Pheidole
JS-C-118	2	9.03	0	ground	FORMICINAE	Gigantiops
JS-C-119	3	8.63	1	ground	MYRMICINAE	Trachymyrmex
JS-C-120	3	10.11	0	ground	ECITONINAE	Eciton
JS-C-121	3	6.69	3	NA	FORMICINAE	Camponotus
JS-C-122	3	11.92	0	ground	ECITONINAE	Eciton
JS-C-123	3	7.87	0	ground	MYRMICINAE	Pheidole
JS-C-124	3	13.62	0	ground	PONERINAE	NA
JS-C-125	3	10.59	0	arboreal	DOLICHODERINAE	Azteca

Table A3.1 (Continued): collections data

Colony Accession	qPCR N	∂15N	Visual Score	Habitat	Subfamily	Genus
JS-C-126	3	11.27	0	arboreal	MYRMICINAE	Crematogaster
JS-C-127	3	12.00	0	ground	MYRMICINAE	Pheidole
JS-C-128	3	13.04	0	ground	ECITONINAE	Labidus
JS-C-130	3	9.46	0	arboreal	FORMICINAE	Nylanderia
JS-C-131	3	7.37	0	ground	MYRMICINAE	Pheidole
JS-C-132	3	7.92	0	ground	MYRMICINAE	Solenopsis
JS-C-136	3	7.80	0	arboreal	MYRMICINAE	Pheidole
JS-C-137	3	9.20	3	arboreal	FORMICINAE	Myrmelachista
JS-C-138	3	11.00	0	ground	MYRMICINAE	Solenopsis
JS-C-140	3	3.60	4	arboreal	FORMICINAE	Camponotus
JS-C-141	3	8.09	0	ground	MYRMICINAE	Atta
JS-C-142	3	6.08	4	arboreal	FORMICINAE	Camponotus
JS-C-143	3	6.38	0	arboreal	MYRMICINAE	Solenopsis
JS-C-145	3	6.67	0	arboreal	MYRMICINAE	Solenopsis
JS-C-147	3	5.94	0	arboreal	DOLICHODERINAE	Azteca
JS-C-148	3	4.27	4	NA	FORMICINAE	Camponotus
JS-C-149	3	7.10	0	arboreal	MYRMICINAE	Solenopsis
JS-C-150	3	6.56	0	arboreal	DOLICHODERINAE	Azteca
JS-C-151	3	4.24	0	ground	MYRMICINAE	Solenopsis
JS-C-152	3	12.76	4	ground	ECITONINAE	Labidus
JS-C-153	3	8.73	0	arboreal	DOLICHODERINAE	Azteca
JS-C-156	2	8.69	0	NA	PONERINAE	Pachycondyla
JS-C-157	3	9.48	0	ground	PONERINAE	Neoponera
JS-C-158	3	10.18	0	ground	PONERINAE	Neoponera
JS-C-159	3	10.55	2	ground	ECTATOMMINAE	Gnamptogenys
JS-C-162	3	8.69	0	arboreal	MYRMICINAE	Crematogaster
JS-C-163	3	9.73	1	arboreal	DOLICHODERINAE	Azteca
JS-C-165	3	11.96	0	ground	MYRMICINAE	Pheidole
JS-C-166	3	5.96	0	NA	DOLICHODERINAE	Dolichoderus
JS-C-167	3	13.72	3	ground	CERAPACHYINAE	Acanthostichus
JS-C-168	3	6.67	0	arboreal	MYRMICINAE	Cephalotes
JS-C-169	3	4.58	0	arboreal	PSEUDOMYRMICINAE	Pseudomyrmex
JS-C-171	3	5.16	0	arboreal	PSEUDOMYRMICINAE	Pseudomyrmex
JS-C-172	3	3.44	0	arboreal	PSEUDOMYRMICINAE	Pseudomyrmex
JS-C-173	3	5.14	3	arboreal	FORMICINAE	Camponotus
JS-C-174	3	3.90	4	arboreal	MYRMICINAE	Cephalotes
JS-C-175	3	5.18	3	NA	DOLICHODERINAE	Dolichoderus
JS-C-176	3	NA	0	ground	PARAPONERINAE	Paraponera
JS-C-177	3	13.45	4	ground	MYRMICINAE	Basiceros
JS-C-178	3	9.45	0	ground	MYRMICINAE	Megalomyrmex
JS-C-179	3	13.71	1	ground	PONERINAE	Odontomachus
JS-C-180	3	6.79	2	arboreal	DOLICHODERINAE	Azteca
JS-C-181	3	6.96	0	arboreal	DOLICHODERINAE	Azteca
JS-C-182	3	5.82	0	arboreal	MYRMICINAE	Allomerus
JS-C-183	3	6.85	0	arboreal	PONERINAE	Odontomachus
JS-C-184	3	3.04	1	arboreal	PONERINAE	Neoponera
JS-C-185	3	6.94	2	arboreal	MYRMICINAE	Daceton
JS-C-186	3	9.84	0	ground	PONERINAE	Neoponera
JS-C-187	3	3.67	4	arboreal	FORMICINAE	Camponotus
JS-C-188	3	5.52	4	arboreal	DOLICHODERINAE	Dolichoderus
JS-C-189	3	4.44	3	arboreal	DOLICHODERINAE	Dolichoderus
JS-C-190	3	6.51	1	arboreal	PSEUDOMYRMICINAE	Pseudomyrmex

Table A3.2: generalized linear mixed model selection

Model	Variable	Fixed effects	Random effects	AIC	*p*
glmm.0	BactPres	~ Habitat	Genus	89.424	
glmm.1	BactPres	~ d15N	Genus	89.587	1
glmm.2	BactPres	~ d15N + Habitat	Genus	91.328	0.61089
glmm.3	**BactPres**	**~ d15N + Habitat + d15N:Habitat**	**Genus**	**83.308**	**0.00155**
glmm.4	BactPres	~ log([DNA]) + d15N + Habitat + d15N:Habitat	Genus	85.3	0.92975

Table A3.3: linear mixed model selection

Model	Variable	Fixed effects	Random effects	AIC	*p*
lmm.0	log(16S)	~ log([DNA])	Colony / Genus	1207.75	
lmm.1	log(16S)	~ log([DNA]) + Habitat	Colony / Genus	1200.01	0.0018
lmm.2	log(16S)	~ log([DNA]) + d15N	Colony / Genus	1200.61	
lmm.3	log(16S)	~ log([DNA]) + d15N + Habitat	Colony / Genus	1199.23	0.066
lmm.4	**log(16S)**	**~ log([DNA]) + d15N + Habitat + d15N:Habitat**	**Colony / Genus**	**1196.36**	**0.0272**

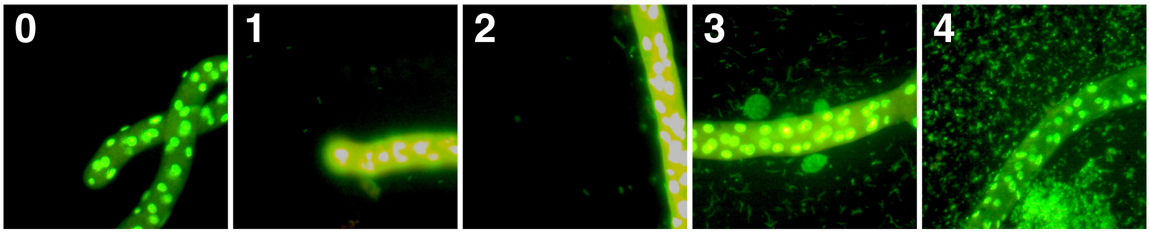


Fig. A3.1: Example micrographs illustrating the visual scale used to estimate bacterial abundances in the field.

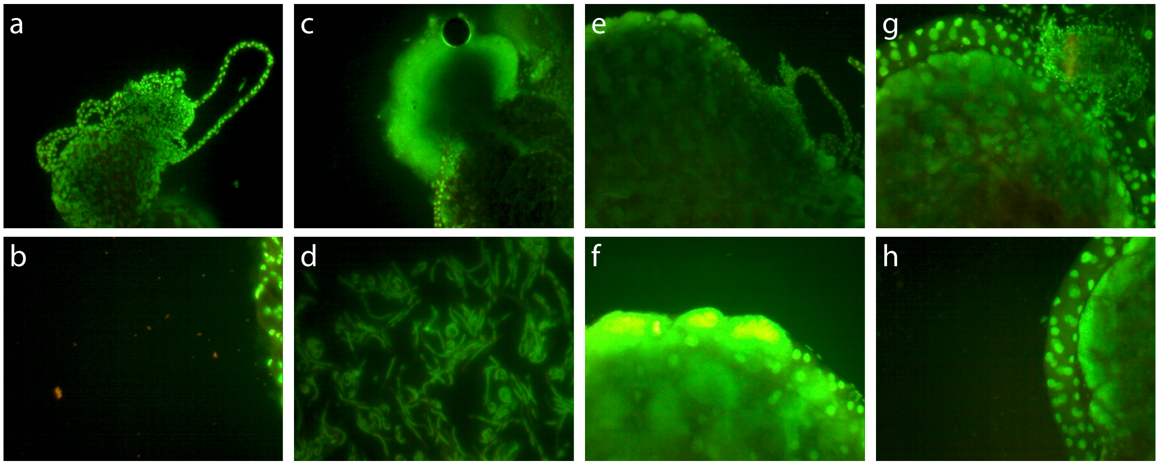


Fig. A3.2: Example fluorescence micrographs taken in the field. All images are uncropped and unmodified. a) An example of an ant gut with no apparent bacteria (IMG-376, Colony 136, *Solenopsis* sp., 4X objective).

b) Visible autofluorescence from gut contents (red) and host cells (green). (IMG-188, Colony 103, *Atta* sp., 10X objective). c) An example of an ant gut with very high apparent bacterial load (IMG-101, Colony 94, *Cephalotes* sp., 4X objective). d) High magnification view of putative *Blochmannia* cells from *Camponotus*. Note stained DNA clumping in apparently contiguous cell bodies, suggesting polyploidy. (IMG-455, Colony 148, *Camponotus* sp., 40X objective) e) Bacteriocytes visible around the midgut of *Camponotus*. (IMG-218, Colony 107, *Camponotus* sp., 4X objective). f) Bacteriocytes visible around the midgut of *Camponotus*. (IMG-219, Colony 107, *Camponotus* sp., 10X objective). g) *Camponotus*-like bacteriocytes visible in *Myrmelachista*. (IMG-227, Colony 108, *Myrmelachista* sp., 10X objective). h) *Camponotus*-like bacteriocytes visible in *Myrmelachista*. (IMG-228, Colony 108, *Myrmelachista* sp., 10X objective)

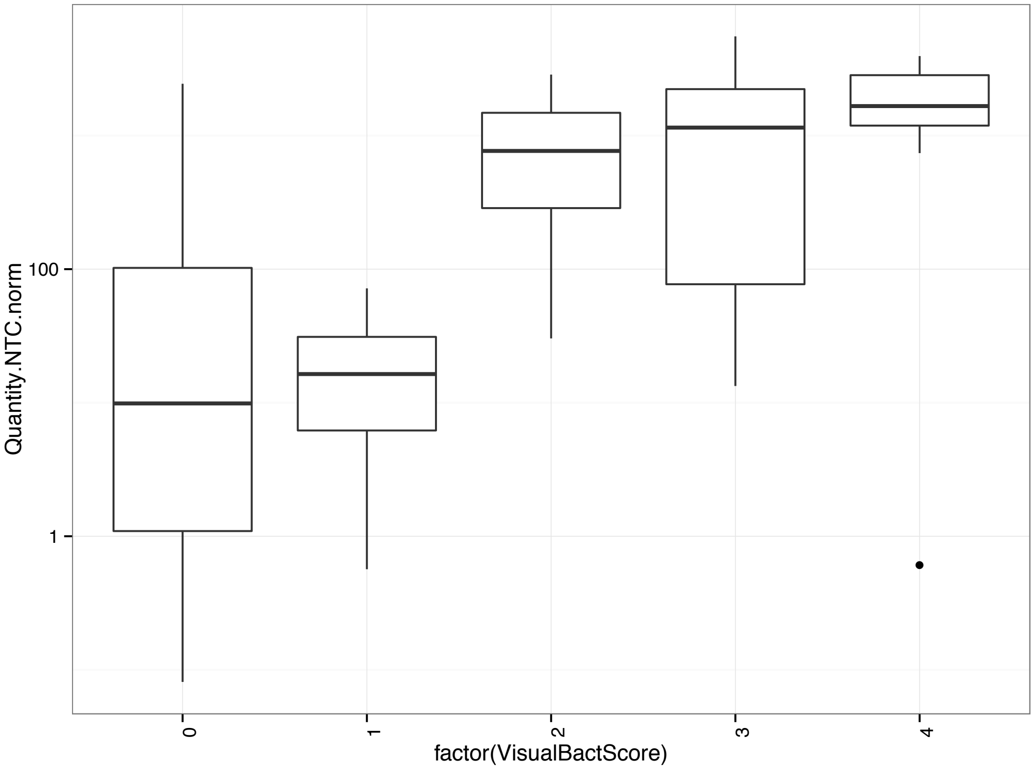
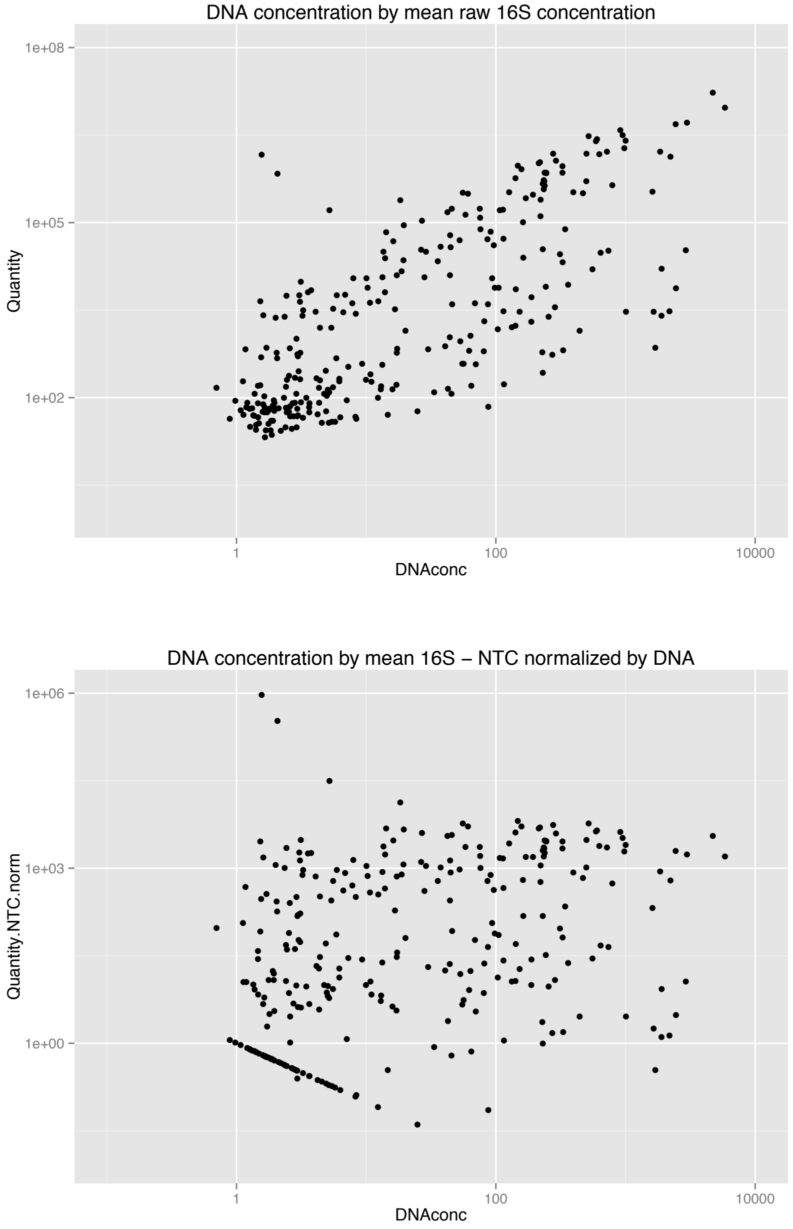


Fig. A3.3: Correlation between visual bacterial abundance estimates and quantitative PCR estimates. qPCR estimates are normalized by DNA concentration and log transformed. Each qPCR data point represents the median value for a single colony (n = 3 workers).


DNA concentration by mean raw 16S concentration

DNA concentration by mean 16S – NTC normalized by DNA

Fig. A3.4: a) Raw estimates of bacterial abundance from quantitative PCR. Each point represents the mean of 2-3 technical replicates from a single worker gut. b) Normalized estimates of bacterial abundance from quantitative PCR. Each value has been reduced by the mean background amplification value (or to a single copy if less than mean background amplification) and divided by DNA concentration.

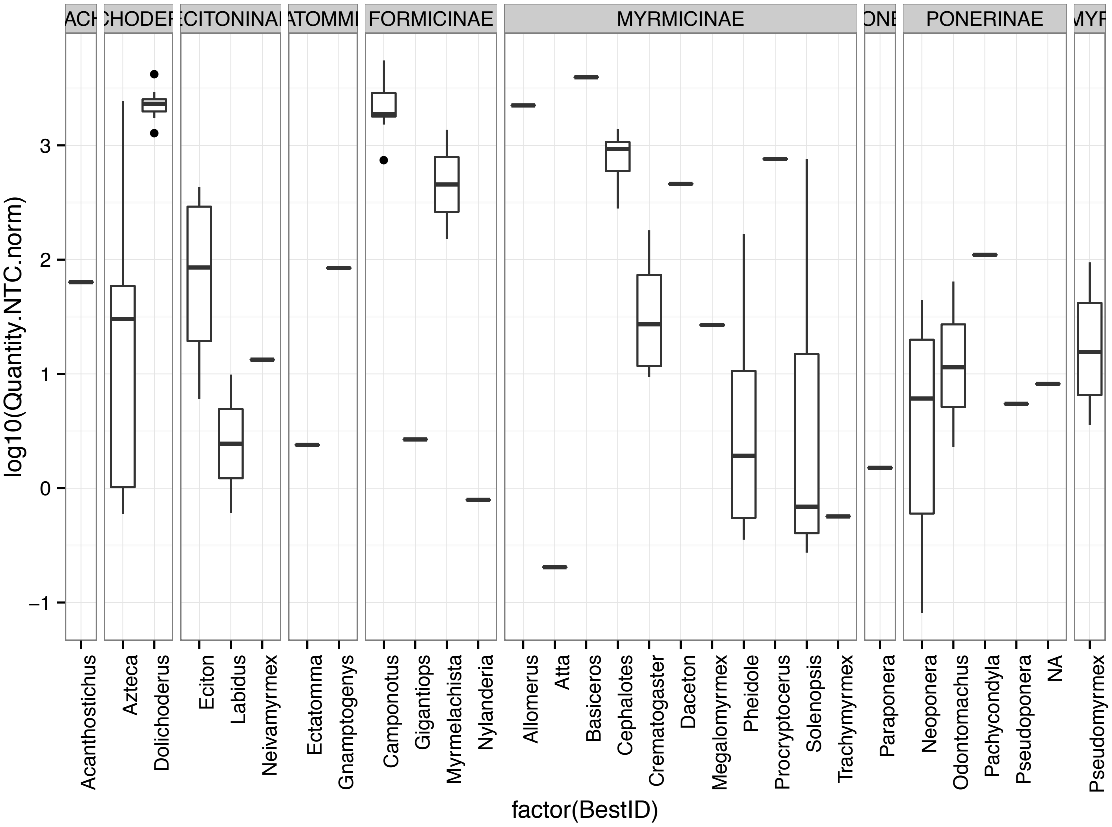


Fig. A3.5: Normalized bacterial abundances by genus. As in Fig. 3, but log-transformed to better display low-range values. Data shown are 16S qPCR counts, minus mean non-template control counts, divided by total DNA concentration. Each data point represents a single colony, taken as the median of three individuals.

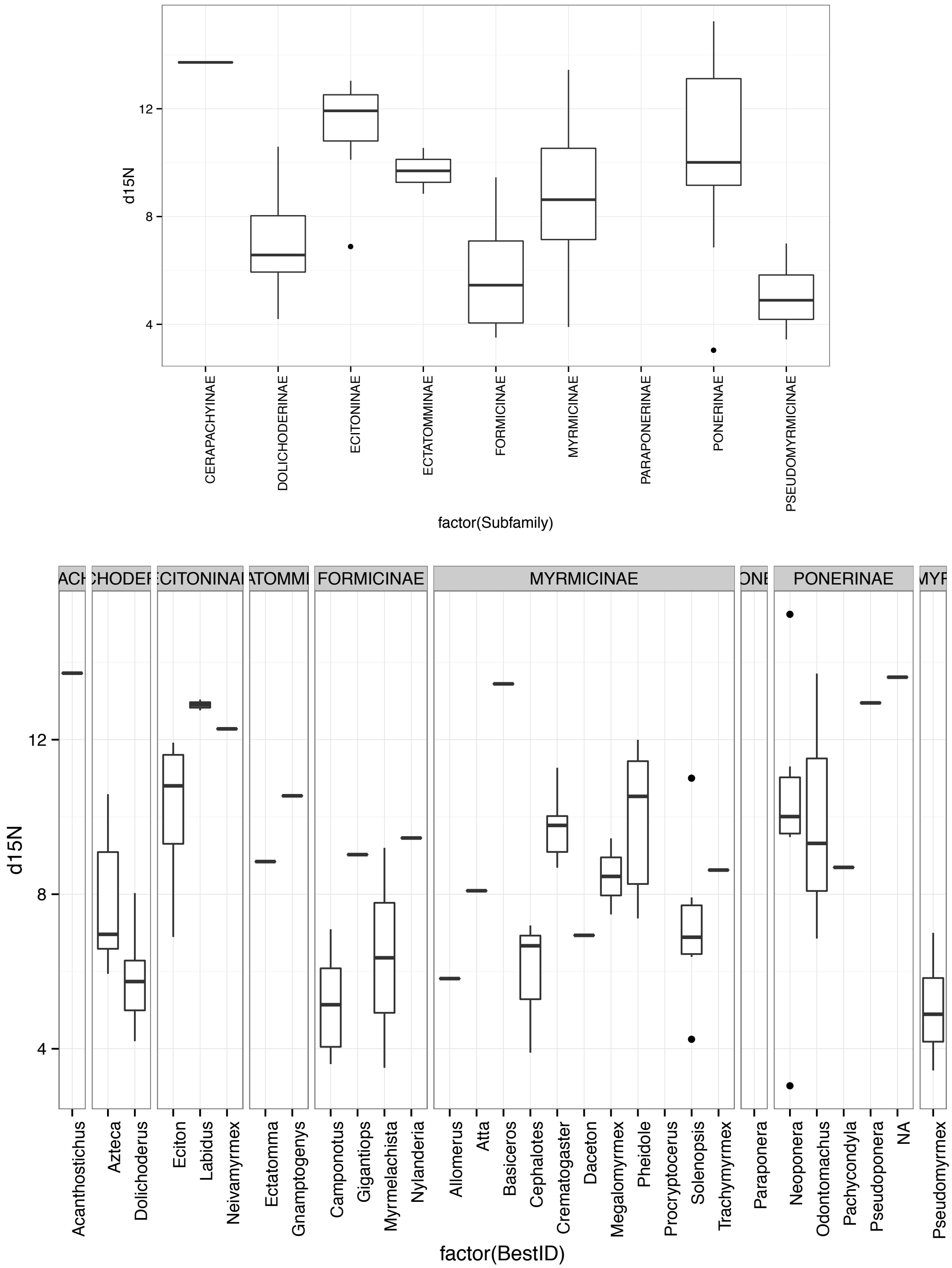


Fig. A3.6: ∂^{15}N isotope ratios. a) Per subfamily. b) Per genus, split by subfamily.

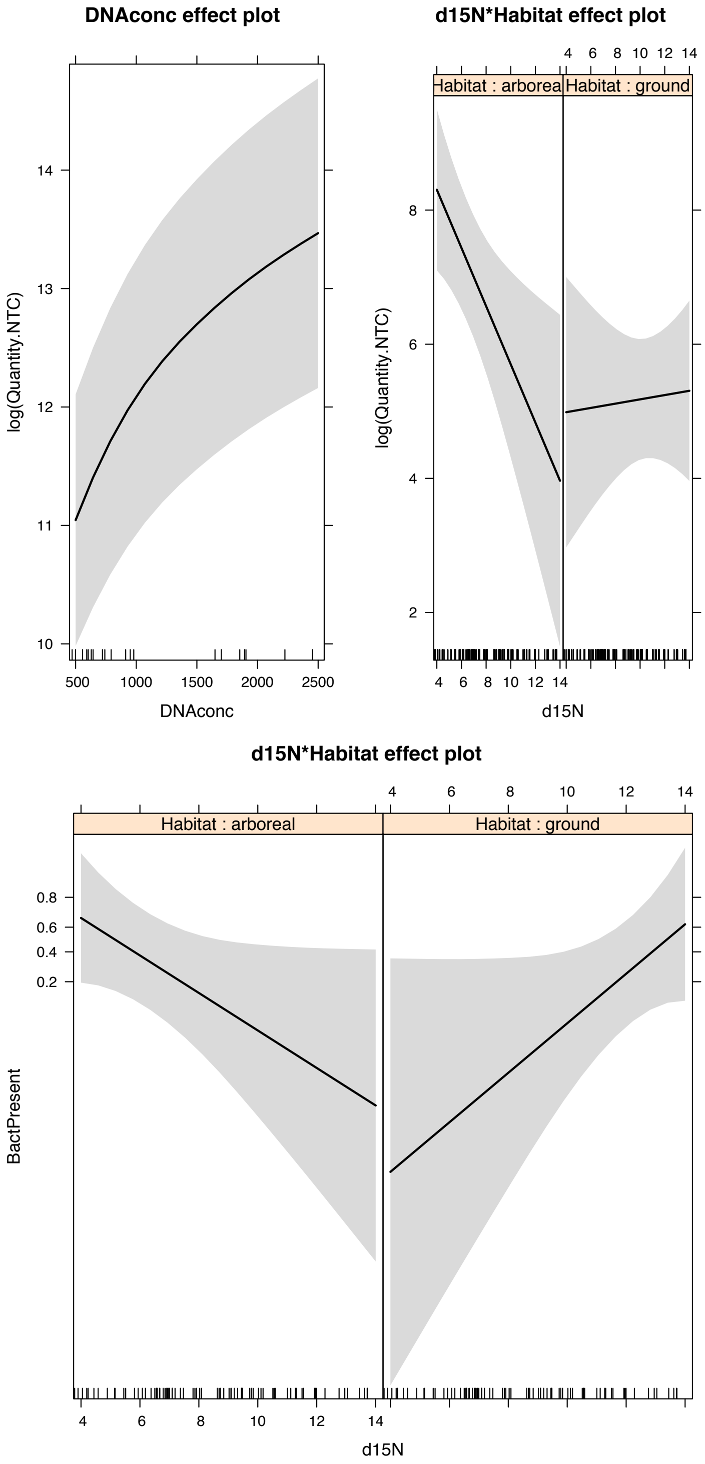


Fig. A3.7: Effect plots for a) mixed-effects linear model of qPCR data and b) generalized linear mixed-effects model of bacterial presence based on visual abundance estimates.

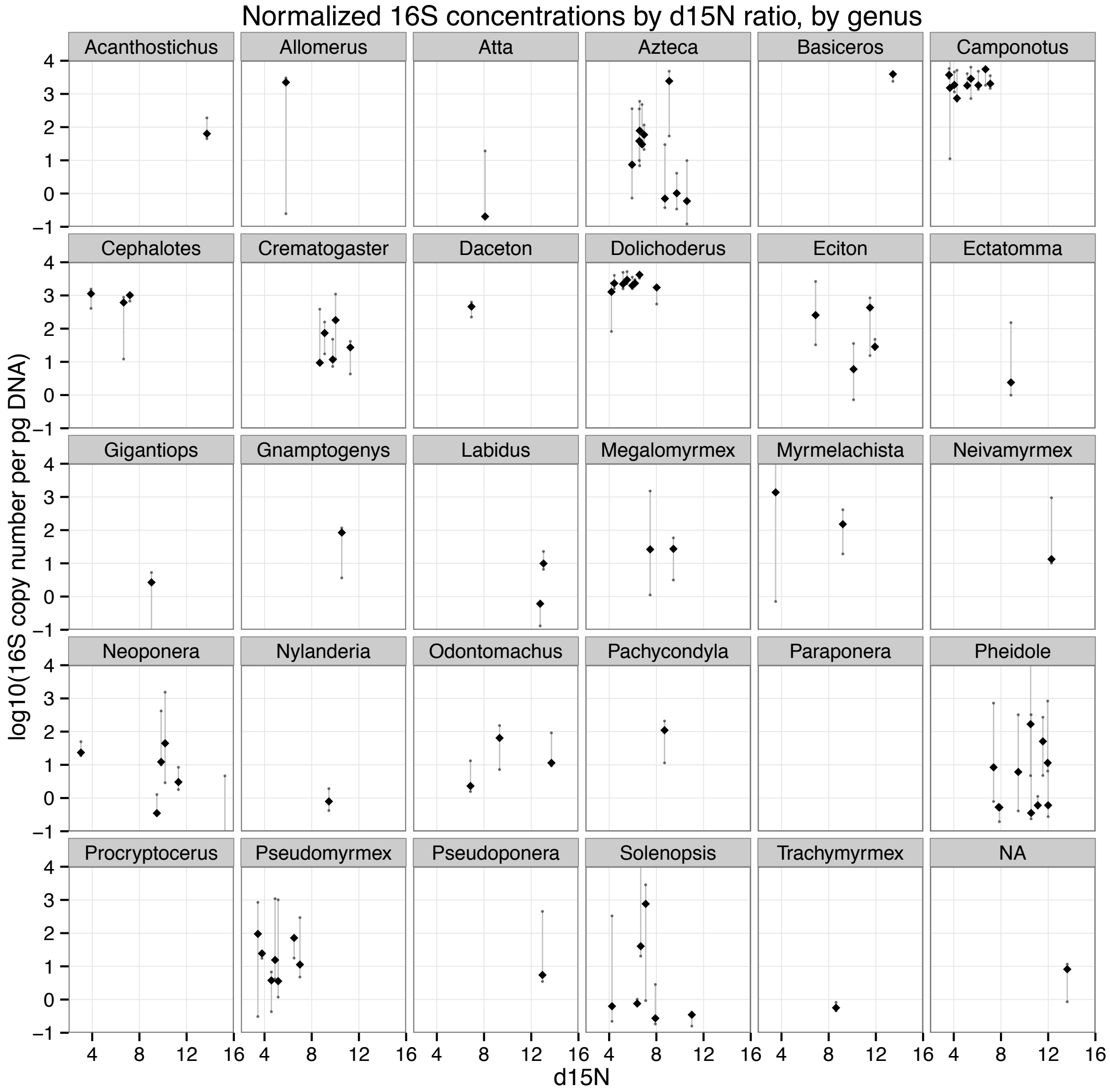


Fig. A3.8: Normalized $\log_{10}$ bacterial 16S copy number by ∂^{15}N isotope ratio, per genus. Large diamonds represent median values per colony, with small points representing individuals. Lines indicate range of values observed for each colony. Each graph is on the same scale.

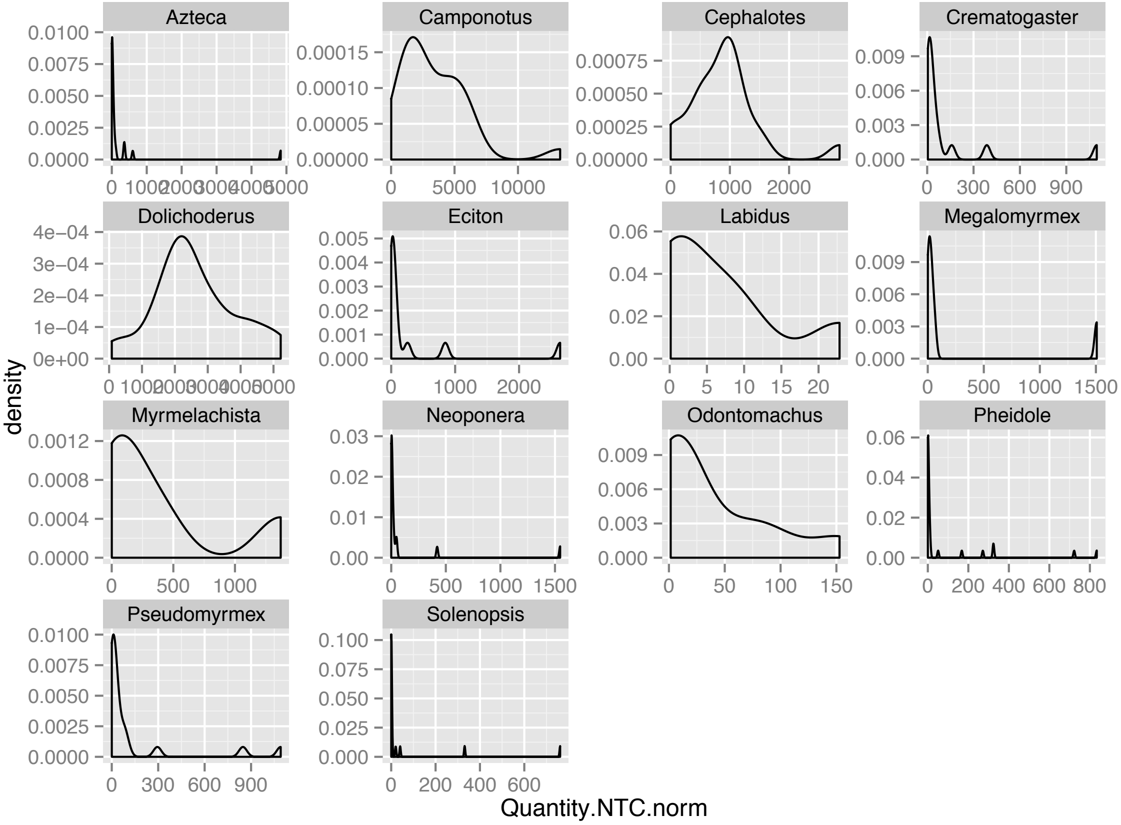


Fig. A3.9: Kernel density plots showing the distribution of normalized 16S concentrations per genus. Distributions represent a pool of all individuals from the genus. Only genera with data from ≥ 6 individuals are represented.

APPENDIX 4

Cladescan: a program for automated phylogenetic sensitivity analysis[2]

Abstract

Examination of trees for the presence of particular nodes is a fundamental aspect of systematics, and is the basis of phylogenetic sensitivity analysis, but becomes unwieldy when performed manually for complex nodes or over large numbers of trees. The program Cladescan is presented here as a stand-alone application to facilitate the detection of nodes in such situations. Cladescan includes features useful for phylogenetic sensitivity analysis, such as automatic generation of 'Navajo Rug' sensitivity plots. Researchers may also find it useful for general comparisons among large data sets.

Introduction

Sensitivity analysis measures how variation in output can be apportioned to variation in input (Saltelli, 2000). In phylogenetics, sensitivity analysis has been used to refer to how relationship hypotheses (phylogenetic trees) vary in response to different tree-search methods (e.g. maximum likelihood, Bayesian likelihood, and maximum parsimony), optimality parameters (e.g. cost matrices in maximum parsimony or models of evolution in maximum likelihood), or character weighting (Wheeler, 1995; Giribet, 2003). As originally proposed by Wheeler (1995), parameter sensitivity analysis was combined with measures of character and taxon congruence among hypotheses as a method for choosing optimal cost parameters for parsimony analysis. However, sensitivity analysis may also be used independently of hypothesis selection to illustrate the stability of a hypothesis to changes in the underlying assumptions, a process implicit in the common practice of displaying both ML and MP bootstrap values on nodes of published trees.

More formally, this leads to the notion of 'nodal stability,' a measure of robustness to input assumptions which can be seen as complementary to measures of nodal support, such as bootstrap values and Bremer support (Giribet, 2003). Support and stability measures are frequently correlated, but when divergent may help to identify nodes of particular interest to the investigator. While this approach has drawn criticism on epistemological grounds (Grant and Kluge, 2003; 2005), researchers may continue to find the technique useful (see D'Haese, 2002; Giribet *et al.*, 2005).

As currently implemented, however, the process of determining nodal stability requires unwieldy manual examination of multiple trees. While this may be trivial for small numbers of taxa and trees, as data sets increase in size and the number of trees grows, manual examination becomes time consuming and prone to human error.

The program 'Cladescan' was written to facilitate such comparisons. Cladescan takes as input one or more sets of trees and a configuration file identifying a node or nodes of interest, records whether each node of interest occurs in each tree, and outputs a summary of the results. This functionality dramatically increases the speed and accuracy of nodal stability estimates compared to manual examination, and may be used as a general tool for quickly comparing topologies among trees.

---

[2] Published: Sanders JG (2010) Program note: Cladescan, a program for automated phylogenetic sensitivity analysis. Cladistics-The International Journal Of The Willi Hennig Society, **26**, 114–116.

Program Description

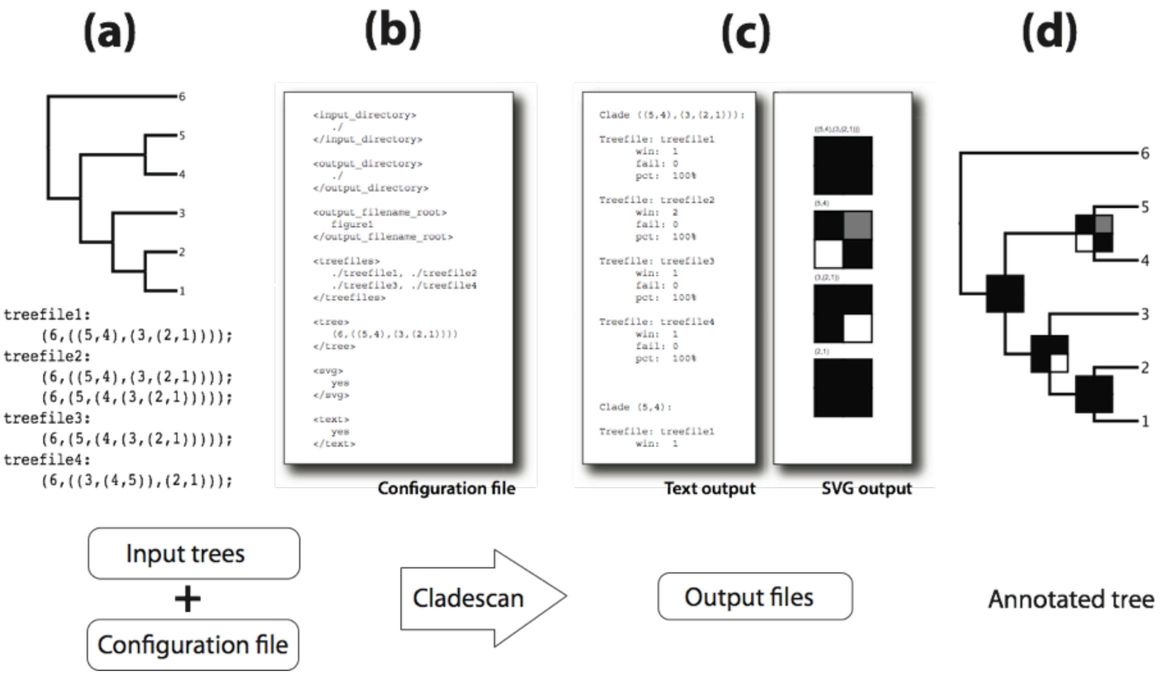
Cladescan is written in Perl, and should run on any Unix-based computer with Perl 5.8.x or higher. The program, along with detailed configuration instructions and sample input files, is available for free download under the GNU General Public License at https://rc.fas.harvard.edu/resources/documentation/software/cladescan/.

Input to Cladescan consists of one or more tree files and one configuration file. Tree files may contain multiple rooted or unrooted trees, which must be in parenthetical format and delimited by semicolons (Figure A4. 1). Each tree file is taken to represent one condition in the analysis (for example, the most parsimonious trees from one particular set of cost parameters); thus, a separate tree file is required for each condition. Output options are specified in the configuration file. The user may indicate a number of specific nodes to annotate, in which case the program will scan for any monophyletic grouping of those terminals; alternatively, one may supply a parenthetical tree, in which case the program will scan for each node of the given tree in turn.

As output, Cladescan produces a text file with the results for each target node and tree file, indicating the number of trees in each file containing each target node, the number not containing the node, and the percentage of trees in that tree file containing the node. Optionally, the program can graphically illustrate these results as ‘Navajo rug’ sensitivity plots (*sensu* Giribet, 2003) in Scalable Vector Graphics (SVG) format, suitable for import into vector-based graphics programs such as Adobe Illustrator (Figure A4.1).

Acknowledgements

The author wishes to thank Ronald Clouse and Gonzalo Giribet for discussion and helpful critiques of this manuscript. This work was supported by a National Science Foundation Graduate Research Fellowship.


(a)
(b)
(c)
(d)
6
5
4
3
2
1
Configuration file
Text output
SVG output
((5,4),(3,(2,1)))
(5,4)
(3,(2,1))
(2,1)
Input trees
+
Configuration file
Cladescan
Output files
Annotated tree

```
treefile1:
    (6,((5,4),(3,(2,1))));
treefile2:
    (6,((5,4),(3,(2,1))));
    (6,(5,(4,(3,(2,1)))));
treefile3:
    (6,(5,(4,(3,(2,1)))));
treefile4:
    (6,((3,(4,5)),(2,1)));
```

```
<input_directory>
   ./
</input_directory>

<output_directory>
   ./
</output_directory>

<output_filename_root>
   figure1
</output_filename_root>

<treefiles>
   ./treefile1, ./treefile2
   ./treefile3, ./treefile4
</treefiles>

<tree>
   (6,((5,4),(3,(2,1))))
</tree>

<svg>
   yes
</svg>

<text>
   yes
</text>
```

```
Clade ((5,4),(3,(2,1))):

Treefile: treefile1
      win:  1
      fail: 0
      pct:  100%

Treefile: treefile2
      win:  2
      fail: 0
      pct:  100%

Treefile: treefile3
      win:  1
      fail: 0
      pct:  100%

Treefile: treefile4
      win:  1
      fail: 0
      pct:  100%



Clade (5,4):

Treefile: treefile1
      win:  1
```

Figure A4.1: Example implementation of Cladescan on a simplified dataset.(a) Maximum-parsimony analysis of a 6-taxon dataset under four different character weightings has resulted in four most-parsimonious tree files, one of which contains two equally parsimonious trees. The investigator wishes to determine whether the nodes from the illustrated tree were recovered in each of the four parsimony condition sets. (b) A configuration file is written, directing the program to the location of the input tree files, specifying which nodes to search for, and indicating the desired output formats. (c) Cladescan searches the input tree files for the specified nodes, then outputs results both as a text file containing detailed information for each node and tree file, and as a graphical representation in SVG format. (d) SVG illustrations can then be manually placed on the tree to illustrate nodal stability across cost matrices.

REFERENCES

Abubucker S, Segata N, Goll J *et al.* (2012) Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Computational Biology*, **8**, e1002358.

Adams ES (1990) Interaction between the ants *Zacryptocerus maculatus* and *Azteca trigona*: Interspecific parasitization of information. *Biotropica*, **22**, 200–206.

Aguirre de Cárcer D, Denman SE, McSweeney C, Morrison M (2011) Evaluation of subsampling-based normalization strategies for tagged high-throughput sequencing data sets from gut microbiomes. *Applied and Environmental Microbiology*, **77**, 8795–8798.

Anderson KE, Russell JA, Moreau CS *et al.* (2012) Highly similar microbial communities are shared among related and trophically similar ant species. *Molecular Ecology*, **21**, 2282–2296.

Archetti M, Scheuring I, Hoffman M *et al.* (2011) Economic game theory for mutualism and cooperation. *Ecology Letters*, **14**, 1300–1312.

Aylward FO, Burnum KE, Scott JJ *et al.* (2012) Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *The ISME Journal*, **6**, 1688–1701.

Beier S, Bertilsson S (2013) Bacterial chitin degradation-mechanisms and ecophysiological strategies. *Frontiers in Microbiology*, **4**, 149.

Billen J, Buschinger A (2000) Morphology and ultrastructure of a specialized bacterial pouch in the digestive tract of Tetraponera ants (Formicidae, Pseudomyrmecinae). *Arthropod Structure & Development*, **29**, 259–266.

Bininda-Emonds ORP, Cardillo M, Jones KE *et al.* (2007) The delayed rise of present-day mammals. *Nature*, **446**, 507–512.

Blatrix R, Djieto-Lordon C, Mondolot L *et al.* (2012) Plant-ants use symbiotic fungi as a food source: new insight into the nutritional ecology of ant-plant interactions. *Proceedings Of The Royal Society B-Biological Sciences*, **279**, 3940–3947.

Blochmann F (1888) Ueber das regelmässige Vorkommen von bakterienähnlichen Gebilden in den Geweben und Eiern verschiedener Insecten. *Zeitschrift fur Biologie*, **24**, 1–676.

Blomberg SP, Garland T, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, **57**, 717–745.

Boettcher K, Ruby E, McFallNgai M (1996) Bioluminescence in the symbiotic squid Euprymna scolopes is controlled by a daily biological rhythm. *Journal Of Comparative Physiology A-Sensory Neural And Behavioral Physiology*, **179**, 65–73.

Breznak JA, Brune A (1994) Role of microorganisms in the digestion of lignocellulose by termites. *Annual Review of Entomology*, **39**, 453–487.

Broderick NA, Buchon N, Lemaitre B (2014) Microbiota-induced changes in drosophila melanogaster host gene expression and gut morphology. *mBio*, **5**, e01117–14.

Brucker RM, Bordenstein SR (2013) The Hologenomic Basis of Speciation: Gut Bacteria Cause Hybrid Lethality in the Genus Nasonia. *Science*, **341**, 667–669.

Buchner P (1965) Endosymbiosis of Animals with Plant Microorganisms (B Mueller, Tran,). John Wiley & Sons, Ltd.

Bution ML, Caetano FH (2008) Ileum of the *Cephalotes* ants: A specialized structure to harbor symbionts microorganisms. *Micron*, **39**, 897–909.

Bution ML, Caetano FH, Zara FJ (2007) Comparative morphology of the ileum of three species of Cephalotes (Formicidae, Myrmicinae). *Sociobiology*, **50**, 355–369.

Byk J, Del-Claro K (2010) Nectar- and pollen-gathering *Cephalotes* ants provide no protection against herbivory: a new manipulative experiment to test ant protective capabilities. *acta ethologica*, **13**, 33–38.

Caetano FH, da Cruz-Landim C (1985) Presence of microorganisms in the alimentary canal of ants of the tribe Cephalotini (Myrmicinae): Location and relationship with intestinal structures. *Naturalia*, **10**, 37–47.

Cantarel BL, Coutinho PM, Rancurel C *et al.* (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, **37**, D233–8.

Cantarel BL, Lombard V, Henrissat B (2012) Complex carbohydrate utilization by the healthy human microbiome. *PLoS ONE*, **7**, e28742.

Caporaso JG, Bittinger K, Bushman FD *et al.* (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, **26**, 266–267.

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

Caporaso JG, Lauber CL, Walters WA *et al.* (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **108**, 4516–4522.

Caporaso JG, Lauber CL, Walters WA *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, **6**, 1621–1624.

Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB (1981) Prokaryotic Cells in the Hydrothermal Vent Tube Worm Riftia pachyptila Jones: Possible Chemoautotrophic Symbionts. *Science*, **213**, 340–342.

Chandler JA, Morgan Lang J, 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.

Colman DR, Toolson EC, Takacs-Vesbach CD (2012) Do diet and taxonomy influence insect gut bacterial communities? *Molecular Ecology*, **21**, 5124–5137.

Cook S, Davidson D (2006) Nutritional and functional biology of exudate-feeding ants. *Entomologia Experimentalis Et Applicata*, **118**, 1–10.

D'Haese CA (2002) Were the first springtails semi-aquatic? A phylogenetic approach by means of 28S rDNA and optimization alignment. *Proceedings of the Royal Society of London B: Biological Sciences*, **269**, 1143–1151.

David LA, Maurice CF, Carmody RN *et al.* (2014) Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, **505**, 559–563.

Davidson D, Patrell-Kim L (1996) Tropical arboreal ants: why so abundant? In: *Neotropical Biodiversity and Conservation* (ed Gibson AC), pp. 127–140. Neotropical Biodiversity and Conservation, Los Angeles.

Davidson DW (2005) Ecological stoichiometry of ants in a New World rain forest. *Oecologia*, **142**, 221–231.

Davidson DW, Cook SC, Snelling RR, Chua TH (2003) Explaining the abundance of ants in lowland tropical rainforest canopies. *Science*, **300**, 969–972.

Davidson D, Cook S, Snelling R (2004) Liquid-feeding performances of ants (Formicidae): ecological and evolutionary implications. *Oecologia*, **139**, 255–266.

de Andrade ML, Baroni Urbani C (1999) *Diversity and adaptation in the ant genus Cephalotes, past and present*. Stuttgarter Beiträge zur Naturkunde Series B (Geologie und Palaëontologie).

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*, **109**, 13034–13039.

Dejean A, Labrière N, Touchard A, Petitclerc F, Roux O (2014) Nesting habits shape feeding preferences and predatory behavior in an ant genus. *Naturwissenschaften*, **101**, 323–330.

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

Desantis TZ, Hugenholtz P, Larsen N *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, **72**, 5069–5072.

Diaz Heijtz R, Wang S, Anuar F *et al.* (2011) Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of Sciences*, **108**, 3047–3052.

Dillon R, Dillon V (2004) The gut bacteria of insects: nonpathogenic Interactions. *Annual Review of Entomology*, **49**, 71–92.

Distel D, Roberts S (1997) Bacterial endosymbionts in the gills of the deep-sea wood-boring bivalves Xylophaga atlantica and Xylophaga washingtona. *Biological Bulletin*, **192**, 253–261.

Douglas A (2006) Phloem-sap feeding by animals: problems and solutions. *Journal Of Experimental Botany*, **57**, 747–754.

Dowd SE, Callaway TR, Wolcott RD *et al.* (2008) Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiology*, **8**, 125.

Dridi B, Fardeau ML, Ollivier B, Raoult D, Drancourt M (2012) *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *International Journal Of Systematic And Evolutionary Microbiology*, **62**, 1902–1907.

Dubilier N, Bergin C, Lott C (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Reviews Microbiology*, **6**, 725–740.

Dussutour A, Simpson SJ (2012) Ant workers die young and colonies collapse when fed a high-protein diet. *Proceedings Of The Royal Society B-Biological Sciences*, **279**, 2402–2408.

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

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996–998.

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

Eilam O, Zarecki R, Oberhardt M *et al.* (2014) Glycan degradation (GlyDeR) analysis predicts mammalian gut microbiota abundance and host diet-specific adaptations. *mBio*, **5**.

Eilmus S, Heil M (2009) Bacterial associates of arboreal ants and their putative functions in an obligate ant-plant mutualism. *Applied and Environmental Microbiology*, **75**, 4324–4332.

Engel P, Moran NA (2013) The gut microbiota of insects - diversity in structure and function. *FEMS Microbiology Reviews*, **37**, 699–735.

Ersts PJ *Geographic Distance Matrix Generator*. American Museum of Natural History, Center for Biodiversity and Conservation.

Ezer A, Matalon E, Jindou S *et al.* (2008) Cell Surface Enzyme Attachment Is Mediated by Family 37 Carbohydrate-Binding Modules, Unique to Ruminococcus albus. *Journal of Bacteriology*, **190**, 8220–8222.

Faith JJ, Guruge JL, Charbonneau M *et al.* (2013) The Long-Term Stability of the Human Gut Microbiota. *Science*, **341**, 1237439–1237439.

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

Feldhaar H, Straka J, Krischke M *et al.* (2007) Nutritional upgrading for omnivorous carpenter ants by the endosymbiont Blochmannia. *BMC Biology*, **5**, 48.

Fernández F, Sandoya S (2004) Syninimic list of Neotropical ants (Hymenoptera: Formicidae). *Biota Colombiana*, **5**, 3–105.

Folgarait PJ, Davidson DW (1995) Myrmecophytic Cecropia - Antiherbivore Defenses Under Different Nutrient Treatments. *Oecologia*, **104**, 189–206.

Frederickson ME, Gordon DM (2007) The devil to pay: a cost of mutualism with Myrmelachista schumanni ants in “devil's gardens” is increased herbivory on Duroia hirsuta trees. *Proceedings of the Royal Society of London B: Biological Sciences*, **274**, 1117–1123.

Frederickson ME, Greene MJ, Gordon DM (2005) Ecology: “Devil's gardens” bedevilled by ants. *Nature*, **437**, 495–496.

Funaro CF, Kronauer DJC, Moreau CS *et al.* (2011) Army ants harbor a host-specific clade of Entomoplasmatales bacteria. *Applied and Environmental Microbiology*, **77**, 346–350.

Gaspar JM, Thomas WK (2013) Assessing the consequences of denoising marker-based metagenomic data. *PLoS ONE*, **8**, e60458.

Gatesy J, Geisler JH, Chang J *et al.* (2013) A phylogenetic blueprint for a modern whale. *Molecular Phylogenetics and Evolution*, **66**, 479–506.

Geib SM, Filley TR, Hatcher PG *et al.* (2008) Lignin degradation in wood-feeding insects. *Proceedings of the National Academy of Sciences*, **105**, 12932–12937.

Gerdts G, Brandt P, Kreisel K *et al.* (2013) The microbiome of North Sea copepods. *Helgoland Marine Research*, **67**, 757–773.

Giribet G (2003) Stability in phylogenetic formulations and its relationship to nodal support. *Systematic Biology*, **52**, 554–564.

Giribet G, Edgecombe GD (2006) Conflict between datasets and phylogeny of centipedes: an analysis based on seven genes and morphology. *Proceedings of the Royal Society of London B: Biological Sciences*, **273**, 531–538.

Gordon DM (2012) The dynamics of foraging trails in the tropical arboreal ant *Cephalotes goniodontus*. *PLoS ONE*, **7**, 1–7.

Grant T, Kluge AG (2003) Data exploration in phylogenetic inference: scientific, heuristic, or neither. *Cladistics-The International Journal Of The Willi Hennig Society*, **19**, 379–418.

Grant T, Kluge AG (2005) Stability, sensitivity, science and heurism. *Cladistics-The International Journal Of The Willi Hennig Society*, **21**, 597–604.

Haas BJ, Gevers D, Earl AM *et al.* (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, **21**, 494–504.

Hackstein JH, van Alen TA (1996) Fecal methanogens and vertebrate evolution. *Evolution*, 559–572.

Hamady M, Knight R (2009) Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research*, **19**, 1141–1152.

Herwig RP, Staley JT, Nerini MK, Braham HW (1984) Baleen whales: preliminary evidence for forestomach microbial fermentation. *Applied and Environmental Microbiology*, **47**, 421–423.

Hölldobler B, Wilson EO (1990) *The Ants*. Harvard University Press, Cambridge.

Hu Y, Łukasik P, Moreau CS, Russell JA (2014) Correlates of gut community composition across an ant species (*Cephalotes varians*) elucidate causes and consequences of symbiotic variability. *Molecular Ecology*, **23**, 1284–1300.

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

Ishak HD, Plowes R, Sen R *et al.* (2011) Bacterial diversity in *Solenopsis invicta* and *Solenopsis geminata* ant colonies characterized by 16S amplicon 454 pyrosequencing. *Microbial Ecology*, **61**, 821–831.

Jaffe K, Caetano F, Sanchez P *et al.* (2001) Sensitivity of ant (*Cephalotes*) colonies and individuals to antibiotics implies feeding symbiosis with gut microorganisms. *Canadian Journal of Microbiology*, **79**, 1120–1124.

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **28**, 27–30.

Kaoutari El A, Armougom F, Gordon JI, Raoult D, Henrissat B (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Reviews Microbiology*, **11**, 497–504.

Kelley ST, Dobler S (2010) Comparative analysis of microbial diversity in *Longitarsus* flea beetles (Coleoptera: Chrysomelidae). *Genetica*, **139**, 541–550.

Kikuchi Y, Hosokawa T, Fukatsu T (2007) Insect-microbe mutualism without vertical transmission: a stinkbug acquires a beneficial gut symbiont from the environment every generation. *Applied and Environmental Microbiology*, **73**, 4308–4316.

Kikuchi Y, Hosokawa T, Nikoh N *et al.* (2009) Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biology*, **7**, 2.

Kitade O (2004) Comparison of symbiotic flagellate faunae between termites and a wood-feeding cockroach of the genus *Cryptocercus*. *Microbes And Environments*, **19**, 215–220.

Knight R, Maxwell P, Birmingham A *et al.* (2007) PyCogent: a toolkit for making sense from sequence. *Genome Biology*, **8**, R171.

Koch H, Schmid-Hempel P (2011) Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proceedings of the National Academy of Sciences*, **108**, 19288–19292.

Kodaman N, Pazos A, Schneider BG *et al.* (2014) Human and Helicobacter pylori coevolution shapes the risk of gastric disease. *Proceedings of the National Academy of Sciences*, **111**, 1455–1460.

Köhler T, Dietrich C, Scheffrahn RH, Brune A (2012) High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Applied and Environmental Microbiology*, **78**, 4691–4701.

Kumar PS, Brooker MR, Dowd SE, Camerlengo T (2011) Target region selection is a critical determinant of community fingerprints generated by 16S pyrosequencing. *PLoS ONE*, **6**, e20956.

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

Langer P (2001) Evidence from the digestive tract on phylogenetic relationships in ungulates and whales. *Journal of Zoological Systematics and Evolutionary Research*, **39**, 77–90.

Lee AH, Husseneder C, Hooper-Bùi L (2008) Culture-independent identification of gut bacteria in fourth-instar red imported fire ant, Solenopsis invicta Buren, larvae. *Journal Of Invertebrate Pathology*, **98**, 20–33.

Lee SM, Donaldson GP, Mikulski Z *et al.* (2013) Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature*, **501**, 426–429.

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

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

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

Longino JT (2006) A taxonomic review of the genus Myrmelachista (Hymenoptera: Formicidae) in Costa Rica. *Zootaxa*, **1141**, 1–54.

Lozupone CA, Knight R (2008) Species divergence and the measurement of microbial diversity. *FEMS Microbiology Reviews*, **32**, 557–578.

Marchler-Bauer A, Lu S, Anderson JB *et al.* (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, **39**, D225–9.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, **17**, pp. 10–12.

Martinson VG, Danforth BN, Minckley RL *et al.* (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular Ecology*, **20**, 619–628.

Martinson VG, Moy J, Moran NA (2012) Establishment of characteristic gut bacteria during development of the honeybee worker. *Applied and Environmental Microbiology*, **78**, 2830–2840.

Matson EG, Ottesen EA, Leadbetter JR (2007) Extracting dna from the gut microbes of the termite (*Zootermopsis angusticollis*) and visualizing gut microbes. *Journal of Visualized Experiments*.

Matsuura Y, Kikuchi Y, Hosokawa T *et al.* (2012) Evolution of symbiotic organs and endosymbionts in lygaeid stinkbugs. *The ISME Journal*, **6**, 397–409.

McCutcheon JP, Moran NA (2011) Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*, **10**, 13–26.

McFall-Ngai M (2007) Adaptive immunity: care for the community. *Nature*, **445**, 153.

McFall-Ngai M, Hadfield MG, Bosch TCG *et al.* (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences*, **110**, 3229–3236.

McNett K, Longino J, Barriga P *et al.* (2009) Stable isotope investigation of a cryptic ant-plant association: *Myrmelachista flavocotea* (Hymenoptera, Formicidae) and *Ocotea* spp. (Lauraceae). *Insectes Sociaux*, **57**, 67–72.

Meredith RW, Janečka JE, Gatesy J *et al.* (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science*, **334**, 521–524.

Meyer F, Paarmann D, D'Souza M *et al.* (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, **9**, 386.

Mira A, Moran NA (2002) Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microbial Ecology*, **44**, 137–143.

Moeller AH, Degnan PH, Pusey AE *et al.* (2012) Chimpanzees and humans harbour compositionally similar gut enterotypes. *Nature Communications*, **3**, 1179.

Moeller AH, Peeters M, Ndjango JB *et al.* (2013) Sympatric chimpanzees and gorillas harbor convergent gut microbial communities. *Genome Research*.

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165–190.

Moreau CS, Bell CD, Vila R, Archibald SB, Pierce NE (2006) Phylogeny of the ants: diversification in the age of angiosperms. *Science*, **312**, 101–104.

Moreau CS, Bell CD (2013) Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution*, **67**, 2240–2257.

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.

Nelson TM, Rogers TL, Brown MV (2013) The gut bacterial community of mammals from marine and terrestrial habitats. *PLoS ONE*, **8**, e83655.

Nelson TM, Rogers TL, Carlini AR, Brown MV (2012) Diet and phylogeny shape the gut microbiota of Antarctic seals: a comparison of wild and captive animals. *Environmental Microbiology*, **15**, 1132–1145.

Nicol S, Hosie GW (1993) Chitin production by krill. *Biochemical Systematics And Ecology*, **21**, 181–184.

Noda S, Kitade O, Inoue T *et al.* (2007) Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotrichonympha* spp.), their hosts, and their bacterial endosymbionts. *Molecular Ecology*, **16**, 1257–1266.

Nyholm SV, Graf J (2012) Knowing your friends: invertebrate innate immunity fosters beneficial bacterial symbioses. *Nature Reviews Microbiology*, **10**, 815–827.

Nyholm SV, McFall-Ngai MJ (2004) The winnowing: Establishing the squid-*Vibrio* symbiosis. *Nature Reviews Microbiology*, **2**, 632–642.

Ochman H, Worobey M, Kuo C-H *et al.* (2010) Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biology*, **8**, e1000546–.

Oksanen J, Blanchet FG, Kindt R *et al.* (2012) vegan: Community Ecology Package.

Oliver KM, Degnan PH, Burke GR, Moran NA (2010) Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, **55**, 247–266.

Olsen MA, Mathiesen SD (1996) Production rates of volatile fatty acids in the minke whale (*Balaenoptera acutorostrata*) forestomach. *The British journal of nutrition*, **75**, 21–31.

Osei-Poku J, Mbogo CM, Palmer WJ, Jiggins FM (2012) Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. *Molecular Ecology*, **21**, 5138–5150.

Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

Paul K, Nonoh JO, Mikulski L, Brune A (2012) "Methanoplasmatales," Thermoplasmatales-related archaea in termite guts and other environments, are the seventh order of methanogens. *Applied and Environmental Microbiology*, **78**, 8245–8253.

Pinto-Tomás AA, Anderson MA, Suen G *et al.* (2009) Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science*, **326**, 1120–1123.

Pope PB, Denman SE, Jones M *et al.* (2010) Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proceedings of the National Academy of Sciences*, **107**, 14793–14798.

Poulsen M, Schwab C, Jensen BB *et al.* (2013) Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nature Communications*, **4**, 1428.

Powell S (2008) Ecological specialization and the evolution of a specialized caste in *Cephalotes* ants. *Functional Ecology*, **22**, 902–911.

Prado-Martinez J, Sudmant PH, Kidd JM *et al.* (2013) Great ape genetic diversity and population history. *Nature*, **499**, 471–475.

Price MN, Dehal PS, Arkin AP (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, **26**, 1641–1650.

Price SL, Powell S, Kronauer DJC *et al.* (2014) Renewed diversification is associated with new ecological opportunity in the Neotropical turtle ants. *Journal of Evolutionary Biology*, **27**, 242–258.

Punta M, Coggill PC, Eberhardt RY *et al.* (2012) The Pfam protein families database. *Nucleic Acids Research*, **40**, D290–301.

Qin J, Li R, Raes J *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464**, 59–U70.

Quince C, Lanzén A, Curtis TP *et al.* (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods*, **6**, 639–641.

Quince C, Lanzén A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **12**, 38.

R Development Core Team R: A Language and Environment for Statistical Computing.

Rahman NA (2015) A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. 1–16.

Rawls JF, Mahowald MA, Ley RE, Gordon JI (2006) Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell*, **127**, 423–433.

Regassa LB, Gasparich GE (2006) Spiroplasmas: evolutionary relationships and biodiversity. *Frontiers in bioscience : a journal and virtual library*, **11**, 2983–3002.

Ren C, Webster P, Finkel SE, Tower J (2007) Increased internal and external bacterial load during Drosophila aging without life-span trade-off. *Cell metabolism*, **6**, 144–152.

Revell LJ (2011) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Roche R, Wheeler D (1997) Morphological specializations of the digestive tract of *Zacryptocerus rohweri* (Hymenoptera: Formicidae). *Journal Of Morphology*, **234**, 253–262.

Roman J, McCarthy JJ (2010) The whale pump: marine mammals enhance primary productivity in a coastal basin. *PLoS ONE*, **5**, e13255.

Roman J, Estes JA, Morissette L *et al.* (2014) Whales as marine ecosystem engineers. *Frontiers in Ecology and the Environment*, **12**, 377–385.

Russell JA, Moreau CS, Goldman-Huertas B *et al.* (2009) Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *PNAS*, **106**, 21236–21241.

Sachs JL, Simms EL (2006) Pathways to mutualism breakdown. *Trends in Ecology & Evolution*, **21**, 585–592.

Sachs JL, Skophammer RG, Regus JU (2011) Evolutionary transitions in bacterial symbiosis. *Proceedings of the National Academy of Sciences*, **108 Suppl 2**, 10800–10807.

Sachs J, Mueller U, Wilcox T, Bull J (2004) The evolution of cooperation. *Quarterly Review Of Biology*, **79**, 135–160.

Salter SJ, Cox MJ, Turek EM *et al.* (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. 1–12.

Sanders JG (2010) Program note: Cladescan, a program for automated phylogenetic sensitivity analysis. *Cladistics-The International Journal Of The Willi Hennig Society*, **26**, 114–116.

Sanders JG, Powell S, Kronaue DJ *et al.* (2013) Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes. *Molecular Ecology*, n/a–n/a.

Savage DC (1977) Microbial ecology of the gastrointestinal tract. *Annual Review Of Microbiology*, **31**, 107–133.

Schmitt-Wagner D, Friedrich MW, Wagner B, Brune A (2003) Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (Cubitermes spp.). *Applied and Environmental Microbiology*, **69**, 6007–6017.

Schoenian I, Spiteller M, Ghaste M *et al.* (2011) Chemical basis of the synergism and antagonism in microbial communities in the nests of leaf-cutting ants. *Proceedings of the National Academy of Sciences*, **108**, 1955–1960.

Segata N, Izard J, Waldron L *et al.* (2011) Metagenomic biomarker discovery and explanation. *Genome Biology*, **12**, R60.

Shannon P, Markiel A, Ozier O *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**, 2498–2504.

Sharon G, Segal D, Ringo JM *et al.* (2010) Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, **107**, 20051–20056.

Shin SC, Kim S-H, You H *et al.* (2011) Drosophila microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science*, **334**, 670–674.

Stoll S, Gadau J, Gross R, Feldhaar H (2007) Bacterial microbiota associated with ants of the genus Tetraponera. *Biological Journal Of The Linnean Society*, **90**, 399–412.

Strøm AR (1979) Biosynthesis of trimethylamine oxide in calanoid copepods: seasonal changes in trimethylamine monooxygenase activity. *Marine Biology*, **51**, 33–40.

Suen G, Scott JJ, Aylward FO *et al.* (2010) An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genetics*, **6**, e1001129.

Sues HD, Reisz RR (1998) Origins and early evolution of herbivory in tetrapods. *Trends in Ecology & Evolution*, **13**, 141–145.

Sun Y, Cai Y, Huse SM *et al.* (2012) A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. *Briefings in bioinformatics*, **13**, 107–121.

Tobin JE (1991) A neotropical rainforest canopy, ant community: some ecological considerations. *Ant-plant interactions*, 536–538.

Turnbaugh PJ, Hamady M, Yatsunenko T *et al.* (2009) A core gut microbiome in obese and lean twins. *Nature*, **457**, 480–U7.

Turnbaugh PJ, Ley RE, Mahowald MA *et al.* (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, **444**, 1027–131.

van Borm S, Buschinger A, Boomsma J, Billen J (2002) *Tetraponera* ants have gut symbionts related to nitrogen-fixing root-nodule bacteria. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, **269**, 2023–2027.

Vijverberg J, Frank TH (1976) The chemical composition and energy contents of copepods and cladocerans in relation to their size. *Freshwater Biology*, **6**, 333–345.

Wang J, Kalyan S, Steck N *et al.* (2015) Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nature Communications*, **6**, 6440.

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.

Weber NA (1957) The nest of an anomalous colony of the arboreal ant *Cephalotes atratus*. *Psyche: A Journal of Entomology*, **64**, 60–69.

Werner JJ, Knights D, Garcia ML *et al.* (2011) Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proceedings of the National Academy of Sciences*, 1–6.

Wheeler DE (1984) Behavior of the ant, *Procryptocerus scabriusculus* (Hymenoptera: Formicidae), with comparisons to other Cephalotines. *Psyche: A Journal of Entomology*, **91**, 171–192.

Wheeler W (1995) Sequence alignment, parameter sensitivity and the phylogenetic analysis molecular-data. *Systematic Biology*, **44**, 321–331.

Williams TM, Haun J, Davis RW, Fuiman LA, Kohin S (2001) A killer appetite: metabolic consequences of carnivory in marine mammals. *Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology*, **129**, 785–796.

Willner D, Daly J, Whiley D *et al.* (2012) Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS ONE*, **7**, e34605.

Wilson EO (1976) A social ethogram of the neotropical arboreal ant *Zacryptocerus varians* (Fr. Smith). *Animal Behaviour*, **24**, 354–363.

Wilson EO (1987) The arboreal ant fauna of Peruvian Amazon forests: a first assessment. *Biotropica*, **19**, 245–251.

Wolschin F, Holldobler B, Gross R, Zientz E (2004) Replication of the endosymbiotic bacterium *Blochmannia floridanus* is correlated with the developmental and reproductive stages of its ant host. *Applied and Environmental Microbiology*, **70**, 4096–4102.

Wong AC-N, Chaston JM, Douglas AE (2013) The inconstant gut microbiota of Drosophila species revealed by 16S rRNA gene analysis. *The ISME Journal*, **7**, 1922–1932.

Xu Q, Morrison M, Nelson KE *et al.* (2004) A novel family of carbohydrate-binding modules identified with Ruminococcus albus proteins. *Febs Letters*, **566**, 11–16.

Yek SH, Mueller UG (2010) The metapleural gland of ants. *Biological Reviews*, **86**, 774–791.

Yin Y, Mao X, Yang J *et al.* (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, **40**, W445–W451.

Yu DW, Davidson DW (1997) Experimental studies of species-specificity in Cecropia-ant relationships. *Ecological Monographs*, **67**, 273–294.

Zhu L, Wu Q, Dai J, Zhang S, Wei F (2011) Evidence of cellulose metabolism by the giant panda gut microbiome. *Proceedings of the National Academy of Sciences*, **108**, 17714–17719.

Zobell CE, Rittenberg SC (1938) The occurrence and characteristics of chitinoclastic bacteria in the sea. *Journal of Bacteriology*, **35**, 275–287.

—. (2011) Microbiology by numbers. *Nature Reviews Microbiology*, **9**, 628.