Convergent Interactions Among Pitcher Plant Microcosms in North America and Southeast Asia

Convergent interactions among pitcher plant microcosms in North America and Southeast Asia

A dissertation presented

by

Leonora Sophia Gumbel Bittleston

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

Biology

Harvard University Cambridge, Massachusetts

April 2016

Dissertation Advisors: Naomi Pierce & Anne Pringle Leonora Sophia Gumbel Bittleston

Convergent interactions among pitcher plant microcosms in North America and Southeast Asia

Abstract

Ecosystems are composed of diverse suites of organisms whose interactions are mediated by both the biotic and abiotic constraints of their environments. The complexity of ecosystems makes them both resilient and difficult to understand. Analyzing the patterns and constraints of biodiversity across different systems can provide insights about the processes governing the formation and maintenance of communities. One analytical tool is convergence, where similarities emerge from different origins.

In this dissertation, I combine conceptual theory with empirical data to explore how natural selection repeatedly favors particular associations among different interacting species. In Chapter 1, I develop the concept of convergent interactions—the independent emergence of multispecies interactions with similar physiological or ecological functions. A convergent interaction framework facilitates prediction of the ecological roles of organisms (including microbes) in multispecies interactions and the selective pressures acting in poorly understood or newly discovered multispecies systems.

The modified leaves of carnivorous pitcher plants are elegant natural systems for studying ecosystem dynamics, as they are discrete, naturally replicated microcosms that have evolved independently three times on different continents. Pitchers house entire

communities of arthropods, protists, fungi and bacteria. In Chapter 2, I take advantage of the visibility and existing knowledge of pitcher plant insect inquilines to evaluate next-generation metabarcoding as a means of characterizing complete contained communities. Correspondence of phylogenetic trees and correlations of organism and sequence counts confirm the effectiveness of metabarcoding methods.

Chapter 3 then extends the concept of convergent interactions to pitcher ecosystems. I characterize and compare the eukaryotic and bacterial communities from over 400 samples of pitcher microcosms from *Nepenthes* species in Southeast Asia and *Sarracenia* species in North America. Data from field collections as well as a relocation experiment are used to investigate whether convergence in form and function of a host extends to associated communities. Pitcher communities contain fewer species than those of surrounding habitats, and phylogenetically related subsets of bacteria and eukaryotes tend to colonize each system. When in a common environment, *Nepenthes* and *Sarracenia* communities converge in composition. The evolved pitcher form appears to strongly affect fundamental aspects of biodiversity within the pitcher ecosystem, including species richness, phylogenetic diversity, and community composition.

Table of Contents

Page	Content
vi	Acknowledgements
1	Introduction
6	Chapter 1: Convergence in multispecies interactions
38	Chapter 2: Metabarcoding as a tool for investigating arthropod diversity in *Nepenthes* pitcher plants
73	Chapter 3: Convergence among the microbiomes of Southeast Asian and North American pitcher plants
108	Appendix A: Dissecting host-associated communities with DNA barcodes
144	Supplementary Material for Chapter 2
152	Supplementary Material for Chapter 3

Acknowledgements

The past six years have been surprisingly pleasant, mostly due to the wonderful people in my life. My PhD has been a struggle of intensive learning, discovery, self-motivation, self-doubt, time management, people management, and teaching. It would not have been possible without my advisors, committee, family, friends and life partner. I am deeply grateful to all of them.

My advisors, Anne Pringle and Naomi Pierce, have both been amazing mentors: always encouraging me to explore, and caring about me as a whole person. They are insightful, imaginative, supportive, and great teachers. I'm thankful for the many hours they have dedicated to meeting with me, and helping with innumerable drafts of abstracts, papers, talks and posters. We have shared international travels, fieldwork, and many lovely meals together. I look forward to life-long friendships with both of them.

Many thanks to my committee members over the years: Aaron Ellison, Chuck Davis, and David Haig. I've enjoyed our discussions, and always appreciate their new perspectives and helpful suggestions.

Collaborators and research assistants make science a fun and interactive enterprise. I appreciate the time and energy contributed by multiple undergraduate students, and in particular Charlie Wolock, Lila Strominger, and Liz Benson. They cheerfully put up with me as an advisor and mentor and during long hours of lab and fieldwork. I am grateful for additional fieldwork support from Kadeem Gilbert and Dan Sternof Beyer in Southeast Asia and Australia.

I am grateful for my helpful collaborators in Southeast Asia including Drs. Bakhtiar Effendi, Kok Gan Chan, Charles Vairappan, and their students, some of whom

include: Kishneth Palaniveloo, Tulasiramanan Ramachandram, and Xin Yue Chan. Thanks to Drs. Li Daiquin and Gregory Jedd for letting me use their lab space in Singapore, and to Shawn Lum, Stuart Davies, Kang Min Ngo, and Su Shiyu for helping me find plants in the Singapore. In the US, Tom Miller provided invaluable suggestions and field support over two different Florida collecting trips.

The data in this dissertation would not exist without the generous funding I received. I was fortunate to be supported by both an NSF Graduate Research Fellowship and an NSF Doctoral Dissertation Improvement Grant, and for the OEB departmental support as well. I am grateful for the MCZ Putnam Expedition Grants that funded two of my expeditions, one to Borneo and one to Australia.

Permits and approvals were a crucial part of my research. In Southeast Asia, I received permits from Singapore National Parks, from the Sabah Biodiversity Centre, and from Maliau Basin Conservation Area. Singapore and Sabah house astounding biodiversity, and were fantastic research sites. I am so lucky to have worked there. In the United States, Harvard Forest graciously allowed me to do experiments on their land, and the Nature Conservancy allowed me to collect pitcher fluid from multiple sites along the Gulf Coast.

Many friends and colleagues provided intellectual and emotional support during my PhD. For teaching me about research and data analysis, I would like to give special thanks to Jon Sanders, Chris Baker and James Crall. Additional thanks to the other graduate students who have helped with writing and analyses over the years, and to my lovely OEB cohort. A big shout-out to the fantastic Guira Woods Vermont crew, including James & Blake Roberts Crall, Maude Baldwin, Fenna Krienen and Ben

Bradlow. Wonderful weekends in the woods with good company gave me the energy and peace of mind to continue working. Also, thank you to so many friends who are not part of the academic world. Non-science conversations, dancing, swimming, hiking, yoga and dinner parties all helped give me perspective and kept me sane.

I am so lucky to have such an awesome, big, supportive family. Thank you to my parents and my many sweet siblings for endless encouragement. And finally, I want to give my deepest appreciation for my partner, Dan Sternof Beyer. I can't imagine a more inspiring, generous, intelligent, and loving person to spend my life with. Thank you for helping with fieldwork, so many presentations, and the big picture ideas.

Introduction

Thousands of interactions among species lead to the complexity of our planet's ecosystems. In order to understand how ecosystems form, and what leads to their stability or volatility, we need to disentangle why and how different species interact. Convergence of multispecies interactions can be used to explore the ways in which natural selection repeatedly favors particular associations among different organisms.

The first chapter develops the concept of convergent interactions, defined as the independent emergence of multispecies interactions with similar physiological or ecological functions. Convergent interactions, as a new framework, relates to both convergent evolution and community convergence, which highlight how selective pressures can shape unrelated organisms or communities in similar ways; however, it explicitly emphasizes interactions among organisms from different trophic levels and often different kingdoms. A focus on convergent interactions clarifies how natural selection repeatedly favors particular kinds of associations among species. Characterizing convergent interactions in a comparative context is likely to facilitate prediction of the ecological roles of organisms (particularly microbes) in multispecies interactions, and selective pressures acting in poorly understood or newly discovered multispecies systems.

We expand on five examples of convergent interactions: bacterial communities in vertebrate guts, ectomycorrhizal symbioses between plants and fungi, bacteria associated with fungus-growing insects, and the food webs in carnivorous pitcher plant microcosms. Bacterial communities in animals with similar gut morphologies and diets converge in composition and functional repertoire, and illustrate how aspects of the communities can be predicted from a few simple factors. Ectomycorrhizae have evolved independently more than 14 times and have a convergent morphology. Recognition of the ectomycorrhizal diagnostic structure provides an

understanding of the plant-fungal metabolic exchange, even if the particular species are undescribed. Antibiotic-producing bacteria have independently formed associations with fungus-growing ants, beetles, and termites, and appear to defend the cultivated fungi against pathogens. This example illustrates how the presence of 'third parties' in a symbiosis can be inferred from convergent interactions. Unrelated lineages of pitcher plants house organisms with similar functional roles in the aquatic pools within their pitchers. The food webs of pitcher plants enable us to explore convergent interactions among entire microecosystems. However, the microbial components of pitcher systems are still not well described.

We selected pitcher plants as our model system to study convergent interactions and in order to understand to what extent convergent interactions are acting in pitcher systems, we needed effective tools to characterize the complete contained communities. In the second chapter, we take advantage of the visibility and existing knowledge of insect inquilines in pitcher plants in order to evaluate metabarcoding methods. We counted dipteran inquilines from a subset of samples, and barcoded morphospecies using the COI barcode and Sanger sequencing in order to compare more traditional methods with an amplicon sequencing approach. Counts of inquiline arthropods were roughly correlated with scaled 18S sequence abundances, indicating that amplicon sequencing is an effective means of gauging community structure. Moreover, phylogenetic trees from COI and 18S were generally congruent, and the taxonomic assignments from both methods were of comparable quality. Results generated by metabarcoding with 18S amplicon sequencing are on par with more traditional barcoding, while also providing additional, useful information regarding microbial organisms.

The three Singaporean *Nepenthes* pitcher plant species all grow together in each of three different collections sites. Therefore we were able to test for specialization of arthropod

inquilines on particular species. Networks of core inquilines and their host species revealed significant specialization of certain arthropod fauna. We extended our analysis of arthropods and their *Nepenthes* host species in section of a paper included here as Appendix A.

The section on inquiline communities in *Nepenthes* pitchers is part of a larger paper titled: "Dissecting host-associated communities with DNA barcodes." Our study compared the factors structuring insect assemblages with those structuring assemblages of obligate parasites of invertebrates, the gregarine protozoa. We hypothesized that community composition of gregarines parasites would follow the same biodiversity patterns as insects; however, we found opposing patterns in the different taxa. A distance-based redundancy analysis showed that variation in insect community structure was better predicted by *Nepenthes* species than by collecting site, while the variation in gregarine community structure was better predicted by site than by pitcher plant species. Insect inquilines appeared to colonize *Nepenthes* pitchers more deterministically than gregarines, with certain organisms selecting specific host plant species, regardless of the location. Conversely, gregarines apparently colonized pitchers more stochastically, exhibiting a stronger correlation with collection site, an effect potentially caused by dispersal limitation.

The third and final chapter compares the convergently evolved *Nepenthes* and *Sarracenia* pitcher plant systems through extensive *in situ* sampling and a relocation experiment. Here, we tested for convergent interactions among the pitcher plant microcosms, asking: does convergence in form and function of a host lead to convergence of the associated communities?

In order to characterize the two systems, we first sequenced the bacterial and eukaryotic communities from over 300 pitchers of *Nepenthes* species from Southeast Asia and *Sarracenia* species from North America. Certain *Nepenthes* species can actively acidify their pitchers, and

pitcher fluids with pH levels below 4 had different bacterial community compositions from those with higher pH levels. The *Sarracenia* species sampled in our study had either short, wide pitchers or tall, tapered pitchers. The differences in form and accompanying changes strongly influenced *Sarracenia* pitcher communities. Both acidity in *Nepenthes* and pitcher form in *Sarracenia* had stronger effects on community structure than did the geographic locations of the pitchers. Thus, aspects of pitcher physiology and/or morphology affected internal communities in each system.

When comparing *Nepenthes* and *Sarracenia* communities to each other, our results show that communities associated with both genera converge in terms of species richness. Relative to surrounding habitats, pitcher communities have fewer species, suggesting that the habitats favor a subset of available species. While the species found in the two systems differ, phylogenetically related subsets of bacteria and eukaryotes are found in each system. Our relocation experiment revealed that when *Nepenthes* were in a North American *Sarracenia* habitat, the community compositions found in these pitchers converged with those of *Sarracenia* pitcher plants. Surprisingly, communities in artificial pitchers (glass tubes) also converged in terms of overall composition; however, aspects of living pitchers are still more similar to each other than to glass tubes. Larvae of pitcher plant mosquitoes colonized foreign *Nepenthes* species, but not artificial pitcher-shaped tubes. The evolved pitcher form appears to strongly affect fundamental aspects of biodiversity within the pitcher, including species richness, phylogenetic diversity, and community composition.

Convergent interactions can be used to better understand selective pressures structuring biodiversity, particularly the vast unseen microbial biodiversity of our planet. Instead of defining and counting species, we can explore compositional and functional similarities of multispecies

interactions and host-associated communities in order to accelerate the mapping of global microbial diversity and the development of conservation strategies.

Chapter 1: Convergence in multispecies interactions

Published in Trends in Ecology & Evolution, April 2016, Vol. 31, No. 4

Authors: Leonora S. Bittleston[1,2], Naomi E. Pierce[1,2], Aaron M. Ellison[1,3] and Anne Pringle[4]

Addresses:
[1]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138 USA

[2]Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA, USA

[3]Harvard Forest, Harvard University, 324 North Main Street, Petersham, MA 01366 USA

[4]Departments of Bacteriology and Botany, University of Wisconsin-Madison, 430 Lincoln Drive, Madison, WI 53706 USA

Corresponding author: Bittleston, L.S. (lbittles@fas.harvard.edu)

Keywords: Community ecology; convergent evolution; microbe; mutualism; symbiosis.

Abstract:

The concepts of convergent evolution and community convergence highlight how selective pressures can shape unrelated organisms or communities in similar ways. We propose a related concept, *convergent interactions*, to describe the independent evolution of multispecies interactions with similar physiological or ecological functions. A focus on convergent interactions clarifies how natural selection repeatedly favors particular kinds of associations among species. Characterizing convergent interactions in a comparative context is likely to facilitate prediction of the ecological roles of organisms (including microbes) in multispecies interactions, and selective pressures acting in poorly understood or newly discovered multispecies systems. We illustrate the concept of convergent interactions with examples: vertebrates and their gut bacteria; ectomycorrhizae; insect-fungal-bacterial interactions; pitcher-plant food webs; and ants and ant-plants.

Convergence in evolution and ecology

The word *convergence* typically describes *convergent evolution*, the independent evolution of similar traits in different lineages resulting from strong selective pressures: "[a]nimals, belonging to two most distinct lines of descent, may readily become adapted to similar conditions, and thus assume a close external resemblance…" (Darwin 1859). Although convergent evolution is primarily a descriptor of morphological features of animals and plants, it can be used to describe microbes and physiological processes as well (e.g., convergent evolution of transcriptional regulation of gene circuits in bacteria and fungi; see Conant and Wagner 2003).

Convergence is also recognized in ecological assemblages, for example in high altitude plant communities of the Andes, Alps, and Himalayas (Humboldt and Bonpland 1805). In fact, the homogeneity of vegetation in geographically distant biomes was discussed early in the history of ecology (Clements 1916, 1936). The resemblance of high altitude plant communities, or whole communities of plants, birds and lizards in the Mediterranean climates of California, Chile, South Africa, and the Mediterranean Basin, are examples of *community convergence*, defined as the physiognomic similarity of assemblages of co-occurring plants or animals resulting from comparable physical and biotic selective pressures (Mooney and Dunn 1970, Cody and Mooney 1978, Samuels and Drake 1997). Community convergence focuses on community structure and functional traits, but does not explicitly investigate interactions among species.

Convergent interactions

We define *convergent interactions* as *the independent emergence of multispecies interactions with similar physiological or ecological functions.* We define *ecological function* as

the role a species plays in an interaction, community or ecosystem; for example, the excretion of essential amino acids by an endosymbiotic bacterium, or the decomposition of dead leaves by an insect detritivore. Our definition of convergent interactions is purposefully broad, and can be used to generate hypotheses about many kinds of ecological relationships. The concept might be especially useful when thinking about symbioses and microbes; for example, the ecology of microbial gut communities in independently evolved herbivores with similar gut morphology, including kangaroos and bighorn sheep (Ley et al. 2008a). Using convergent interactions as a framework for studying associations is likely to bring new clarity to nascent and dynamic studies of symbioses among microbes and other organisms (reviewed in McFall-Ngai et al. 2013).

Convergent interactions are often associated with convergently evolved morphological structures, and specialized morphologies can aid in identification of ecological functions. For example, any fungus forming a "Hartig net" within a root tip is likely participating in an ectomycorrhizal mutualism with a plant (Smith and Read 2010). Moreover, research on convergent interactions in one location can illuminate similar reciprocal selective pressures acting in analogous systems. For example, experiments with ants and ant-plants in Africa are likely to inform understanding of independently evolved ants and ant-plants in South America or Asia, not unlike using a "prior" in Bayesian inference.

We suggest that explicitly recognizing convergent interactions will provide a heuristic method to predict: 1) the functions of multiple, associated species, such as the metabolic capacities of microbes in an herbivore gut; 2) the ecological role of a symbiosis involving newly discovered or poorly described species, such as an ectomycorrhizal symbiosis recently found in a tropical habitat; and 3) selective pressures acting in one system based on data from a different system, such as among ant-plants found on different continents. Although convergence of

multispecies interactions has been implicitly discussed in recent papers (e.g., Muegge et al. 2011, Fan et al. 2012, Aylward et al. 2014), the concept has never before been explicitly defined or formally explored.

Situating convergent interactions

Our use of *convergent interactions* differs from current uses of convergent evolution and community convergence. *Convergent evolution* is defined strictly by phylogeny, and concerns individual species, not interactions. By contrast, convergent interactions focuses on the ecology and behaviors of multiple interacting species; moreover, the independent evolution of *all* of the interacting species is not required. For example, symbioses of two oak species with distantly related and independently evolved lineages of ectomycorrhizal fungi (e.g. truffles and boletes) can still be considered convergent even though the capacity of the oaks to form ectomycorrhizal associations is a synapomorphy: the associated fungi evolved the ectomycorrhizal habit independently. *Community convergence* describes similarities in the distribution, diversity, and morphologies of geographically disparate sets of co-occurring species in relation to similarities of their habitats (e.g., shrubs or lizards from California and Chile (Mooney and Dunn 1970, Cody and Mooney 1978)), but does not specifically address interspecific interactions. Typically, community convergence focuses on a particular guild. By contrast, convergent interactions emphasizes relationships among multiple organisms and trophic levels, and often across different kingdoms. In certain circumstances, convergent interactions can appear as community convergence, for example if the mammalian gut is defined purely as a habitat and not as part of an organism. However, we think that an association among living organisms (e.g., bacteria and a human gut) will have fundamentally different evolutionary dynamics than an association of

organisms with an abiotic environment (e.g., bacteria and a sewer pipe), because of the potential for coevolution.

A different framework, the geographic mosaic of coevolution, is useful for understanding how natural selection and coevolutionary processes differ among populations (Thompson 2005). By contrast, convergent interactions encompasses interactions among groups of different species emerging independently from different lineages and in different regions of the world (e.g., ant-plant interactions in Africa, Southeast Asia, and South America). Convergent interactions takes a broader perspective than the geographic mosaic of coevolution, because it compares different groups of species across regions; however, geographic mosaics likely act within each group of species in a region (e.g., within Kenyan ant-plant interactions).

Last, convergent interactions also differs from analyses of phylogenetic community ecology. Phylogenetic community ecology examines how species are phylogenetically and phenotypically clustered or over-dispersed within a community, and typically explores whether these patterns are caused by competition or evolutionary convergence of similar traits (Cavender-Bares et al. 2009). Analyses of phylogenetic community ecology normally are done within one phylogenetic clade, and most often within a particular region; oak trees in Florida are a salient example (Bares et al. 2004). By contrast, an analysis of convergent interactions might compare the interspecific interactions of organisms of different phylogenetic clades across geographically distant systems; for example, among trees and ectomycorrhizal symbionts from North America and Australia (Figure 1.1A).

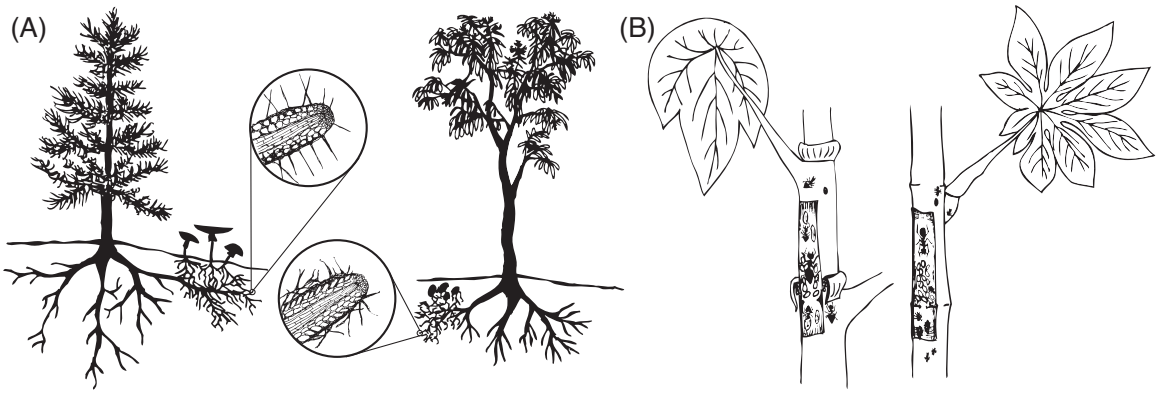


Figure 1.1: Convergent interactions in ectomycorrhizae and ant-plants

Legend: A) Examples of ectomycorrhizal symbioses. Left: A pine tree (*Pinus* [a gymnosperm]) and the basidiomycete *Amanita*. Right: Southern Beech (*Nothofagus* [an angiosperm]) and the ascomycete *Elaphomyces*. Though neither trees nor fungi are closely related, the root tips (enlarged in circles) of both symbioses have similar morphologies: a mantle of fungal hyphae covering the root, highly branched structures between root cells (the Hartig net), and extra-radical mycelia extending from the mantles into surrounding soils. B) Examples of ant-plants and plant-ants. Left: *Macaranga* (Malpighiales) and *Crematogaster* (Myrmicinae). Right: *Cecropia* (Rosales) and *Azteca* (Dolichoderinae). Though neither the plants nor ants are closely related, both trees have domatia in hollow stems where ants rear larvae and tend hemipterans, and both produce food bodies consumed by ants. In turn, the ants protect their trees from herbivores. Illustrations by L. S. Bittleston.

Convergent evolution is best understood in a phylogenetic framework, where trait evolution can be traced through ancestral nodes and the independence of a particular trait can be explored (Stayton 2008, Losos 2011). However, the ancestral nodes of entire communities (however defined) cannot be modeled with phylogenies, because of the continual exchange of species among habitats within a region (Schluter and Ricklefs 1993). Nor can interactions be modeled with phylogenies, unless the interactions involve hosts and symbionts with strict vertical transmission (Funk et al. 2000). Methods for assessing convergent evolution or community convergence can only be applied to convergent interactions when one set of interacting partners are reduced to continuous traits (Boxes 1 and 2). Network analyses with dynamic models that allow for coevolution among interacting species might be useful avenues for developing future methods to assess convergent interactions (Box 2).

Examples of convergent interactions

To develop the concept of convergent interactions, we discuss five examples involving microbial symbioses, mutualisms, and trophic interactions. The examples illustrate how convergent interactions can be used to predict species' functions, ecological relationships, and selective pressures in novel systems. Because we are interested in convergence, we focus on similarities among the interactions; however, documenting differences will also be useful, particularly for understanding how phylogenetic constraints act within evolutionary lineages (Box 3). We note that in any discussion of convergence, "similarity" requires clear definition (Samuels and Drake 1997). At fine scales, for example when species are identified, communities might appear to be random assemblages (Gleason 1926), but at coarser scales, for example when functional groups are identified, convergent patterns emerge (Fukami et al. 2005). We identify the appropriate scale and measure of similarity for each example.

Our first three examples focus on associations between microorganisms and animals or plants. Molecular techniques have greatly increased knowledge of microbial diversity and functions. For example, we now know that gut bacteria influence human weight, nutrition, and immune function (Kau et al. 2011); bacterial and yeast endosymbionts provide insects with nutrition and protection (reviewed in Douglas 2015); and fungal endophytes deter plant pathogens and herbivores (Clay 1988, Arnold et al. 2003, Bittleston et al. 2011). However, the majority of the world's microbes—and their interactions and functions—remain difficult to characterize. A focus on convergent interactions will generate hypotheses to explain observed patterns of microbial species diversity and function.

Bacterial communities in vertebrate guts illustrate how convergent interactions can be used to estimate community composition and functional repertoire based on the morphology and diet of an animal. Gut bacteria living within the digestive systems of animals assist in digestion of complex carbohydrates, provide vitamins, detoxify compounds, facilitate the maturation of the vertebrate immune system, and protect against some pathogens by interfering with other microbes (Walter et al. 2011). Among vertebrates, mammals (especially herbivorous ruminants) have particularly dense and diverse communities of gut microbes. The ancestors of mammals were carnivorous (Stevens and Hume 1995), but the herbivorous habit is extremely successful: 80% of mammals alive today are herbivores, with herbivores present in 11 of 20 mammalian orders (Stevens and Hume 1995).

Communities of gut bacteria in mammals differ from free-living microbial communities, and reflect phylogenetic history, morphology, and host diet (Ley et al. 2008b, 2008a). Among herbivorous mammals, gut morphology is correlated with fecal microbiota composition: foregut and hindgut fermenters have different microbial communities, regardless of the hosts' evolutionary relationships (Ley et al. 2008b). While foregut fermentation evolved separately in ungulates, rodents, marsupials, primates, and birds (Stevens and Hume 1995), the bacterial gut community of the hoatzin (the only avian foregut fermenter) is more similar to the gut community of a cow than it is to that of a chicken (Godoy-Vitorino et al. 2011). Ant and termite-eating vertebrates similarly show convergence of gut microbial communities; in this case, diet is probably the major influence (Delsuc et al. 2014). Gut microbiomes are strongly influenced by both gut morphology and host diet, and distantly related hosts with similar diets tend to independently acquire organisms from the same bacterial phyla (Ley et al. 2008b).

Does relatedness translate to function? The functional repertoires of fecal bacteria isolated from herbivores and carnivores can be predicted from phylogenetic measurements of bacterial species assemblages (Muegge et al. 2011). Gut microbiota of herbivores predominantly produce enzymes for amino acid biosynthesis, whereas gut microbiota of carnivores produce more enzymes involved in amino acid degradation. Bacteria from herbivores, even independently evolved herbivores, generally build amino acids; conversely, bacteria from carnivores generally break down proteins (Muegge et al. 2011). Host diet appears to cause convergence of function, as well as identity.

Convergent interactions are a better descriptor of these relationships than community convergence. Analysis with community convergence would omit the functional nature of the interactions while focusing on environmental filtering of otherwise randomly assembled groups of species.

Ectomycorrhizae illustrate how morphologies can be used to identify convergent interactions, even when the associations involve undescribed species. Mycorrhizae are symbioses between fungi and plants: the fungi supply scarce resources to plants in exchange for carbon (Smith and Read 2010). Ectomycorrhizal (ECM) associations involve both ascomycete and basidiomycete fungi, and are found in every terrestrial ecosystem.

ECM symbioses evolved as recently as 50 Mya (LePage et al. 1997, Beimforde et al. 2011). The associations evolved repeatedly and independently in the ascomycetes and basidiomycetes (Smith and Read 2010); in the latter, ECM associations evolved at least 14 times in at least eight orders (Box 1). Among these 14 origins, eight of the *de novo* symbioses involved angiosperms, whereas six involved gymnosperms (Hibbett and Matheny 2009).

In spite of their independent origins, and in line with the concept of convergent interactions, ECM morphology is convergent across lineages (Figure 1.1A). ECM associations are defined by three features: 1) a fungal sheath or mantle around a root; 2) a network of hyphae (the Hartig net) within the epidermal and cortical cells of the root; and 3) a mycelium extending from the root through soil (Smith and Read 2010). Once the diagnostic morphologies of ectomycorrhizae are recognized, broad aspects of the plant-fungal metabolic exchange are clear, even if the particular species are undescribed. Thus, recognition of ECM morphologies facilitates discovery of mutualisms involving undescribed species of fungi, especially where biodiversity is poorly characterized (e.g. Peay et al. 2009).

Insects that grow fungi and use antibiotic-producing bacteria to defend their gardens from antagonistic organisms illustrate that presence of "third parties" in a symbiosis can be inferred from convergent interactions. Species of ants, beetles, termites, and gall midges all grow fungi as a food source in enclosed "garden" chambers. Associations of insects and fungi have evolved repeatedly and independently, across different orders of both groups (Mueller et al. 2001). In these convergent interactions, fungi provide nutrition and a nesting substrate for the insects, while insects provide material for fungal decomposition, protected growing spaces, and transportation to new locations.

Antibiotic-producing bacteria are a common third party associated with fungus-growing ants, beetles, and termites. For example, ants have developed associations with Actinobacteria multiple times, with independent acquisitions in two genera (Cafaro and Currie 2005, Barke et al. 2010). The Actinobacteria are maintained on the ants' cuticles, and target the garden parasite *Escovopsis* (Cafaro and Currie 2005, Currie et al. 2006, Barke et al. 2010). Two species of bark

beetle also associate with Actinobacteria to suppress antagonistic fungi (Cardoza et al. 2006, Scott et al. 2008). A different bacterial lineage appears to play the same role within termite symbioses: a *Bacillus* species in the phylum Firmicutes appears to selectively target fungi antagonistic to the fungal cultivar farmed by the termite *Macrotermes natalensis* (Um et al. 2013). The functional gene profiles of the bacteria associated with fungus-growing ants, beetles, and termites are convergent (Aylward et al. 2014), with roughly equivalent physiological potentials, even though the insects are from three different orders, and the fungal cultivars are from two different phyla (Aylward et al. 2014) (Box 1).

These independently evolved, close associations of fungus-growing insects and antifungal-producing bacteria provide a powerful model that can be used to understand other symbioses. In contrast to the symbioses involving ants, beetles, and termites, the fungus-growing habit of the gall midges (from the tribes Lasiopterini and Asphondyliini of the family Cecidomyiidae) is poorly understood. Gall midges associate with fungi in what are thought to be obligate nutritional symbioses (Rohfritsch 2008, Heath and Stireman 2010), but gall midges and their associated fungi are rarely been studied. Nonetheless, convergent interactions among bacteria, fungi, and ants, beetles, or termites suggests that selective antibiotic-producing bacteria also will be found in the fungal gardens of gall midges. In these systems, convergence is defined as hosting bacterial taxa with a common function: the production of antimicrobial compounds capable of protecting a target cultivar.

Food webs in carnivorous pitcher plants enable us to explore convergent interactions among entire micro-ecosystems. In three unrelated plant families on three different continents, the pitcher-shaped carnivorous organs are formed from a single, modified leaf (Arber 1941,

Albert et al. 1992). Pitcher plants use extrafloral nectaries to attract insect prey. The pitchers have slippery interior surfaces, and enzymes within pitchers digest prey to access resources that otherwise are scarce in the low nutrient soils where these species grow. Once the pitchers open, food webs of insects, arachnids, protozoa, rotifers, bacteria, and fungi form in the water-filled pools of many pitcher plant species (Frank 1983).

Diverse organisms live in the pitchers; although many are host-specific (Kitching 2000), functions are often similar. For example, both *Sarracenia* (Ericales: Sarraceniaceae) and *Nepenthes* (Caryophyllales: Nepenthaceae) pitchers host predators, filter feeders, and detritivores (Adlassnig et al. 2011). The food webs of pitchers from these different families on different continents are more like each other than they are like the food webs of other aquatic microcosms, even if the microcosms are in the same habitat as the pitchers (Kitching 2000). Convergence might be influenced by the plants' internal chemistry, which is similar among all pitcher plant lineages and is controlled to some extent by the plant (Bradshaw and Creelman 1984, An et al. 2001, Adlassnig et al. 2011).

Pitchers are elegant models that can be used to test for convergent interactions among microecosystems. For example, knowing that predators, filter feeders, and detritivores are common to well-studied *Sarracenia* and *Nepenthes* species suggests hypotheses about the presence, absence, and ecological relationships of communities not only within newly described or poorly studied pitcher plant species, but also in other phytotelms (Ellison et al. 2003, Srivastava et al. 2004). Microbes actively decompose captured prey in *Sarracenia purpurea* pitchers (Bradshaw and Creelman 1984, Butler et al. 2008), likely increasing available resources, and we hypothesize that microbes with similar functional repertoires will be active in the microbiomes of other *Sarracenia* and *Nepenthes* species, as well as in pitchers of the rarely

studied pitcher-plant *Cephalotus follicularis*. As with the insect-fungal-bacterial associations, convergent interactions in these systems are defined by the presence or absence of species with a specific function in the food web; like the bacterial communities of herbivore guts, similarity can be measured using metabolic capacities, such as microbial enzymes for protein decomposition.

Ant-plant mutualisms result from similar selective pressures**.** Ants have associated intimately with plants since at least the diversification of flowering plants almost 100 Mya (Wilson and Hölldobler 2005, Moreau et al. 2006). Different groups of plants have evolved specialized interactions with ants, providing food and nest sites in exchange for protection from herbivores, pathogens, and competing plants (Davidson and McKey 1993). More than 25% of all plant families secrete extra-floral nectar, and plants from at least 20 different families produce hollow thorns or stems ("domatia") that provide ants with suitable nesting sites; many also are provisioned with food bodies rich in protein or fat (Bronstein 2006).

Classic examples of ant-plants include African *Vachellia* (*Acacia*) species, which grow large swollen thorns to house ant colonies (Young et al. 1997); Southeast Asian *Macaranga* species, which host ants within hollow swollen stems (Fiala et al. 1989, Quek et al. 2007); and Neotropical *Acacia* and *Cecropia* species, which have either swollen thorns similar to African *Vachellia* (*Acacia*) or hollow stems similar to *Macaranga* (Janzen 1969) (Figure 1.1B, Box 1). Ant-plants tend to grow quickly and in high-light environments, and the associated ants tend to be aggressive towards other organisms, even removing nearby vegetation to enable the host plant to compete more effectively for light and space (Rico-Gray and Oliveira 2007). As with our example of convergent interactions among ectomycorrhizae, morphology signals a particular

kind of ant-plant interaction; a newly discovered species of plant with swollen hollow stems, food bodies, and extra-floral nectaries is very likely to be in a long-term association with ants.

Because the benefits and costs of ant-plant mutualisms have been demonstrated experimentally in some systems, selective pressures shaping the evolution of ant-plants that have not yet been investigated can be inferred (Rosumek et al. 2009). For example, when the plant-ants *Pseudomyrmex ferruginea* are present on *Vachellia* (*Acacia*) *cornigera*, the plants experience decreased herbivory and increased survival while the ants gain nutrition and nesting space (Janzen 1966). Ant inhabitants compete fiercely for control of domatia and easy access to food provided by ant-plants (Davidson and McKey 1993). Hosting ants can be costly for plants, because of the resources devoted to producing extra-floral nectar and shelter structures. Protecting plants also can be costly for ants. Some ants attack herbivores such as elephants that they cannot kill and eat (Mayer et al. 2014). In spite of these costs, the repeated evolution of ant-plant relationships in all tropical regions of the world likely is caused by the mutual strong selective pressures of herbivory on plants, and competition for food and nesting space on ants (Davidson and McKey 1993).

Mechanisms mediating the emergence of convergent interactions

Why and how do convergent interactions emerge? No organism lives in isolation, and most organisms require both their own genetic information and functions provided by other species to survive and reproduce (Thompson 1999). Phylogenetic constraints can limit potential trait space. In some circumstances, evolving a close interaction with another organism will be simpler than evolving a new metabolic function. For example, prokaryotes are the only organisms able to make nitrogenases, enabling them to fix atmospheric nitrogen (Raymond et al.

2004). Plants are generally nitrogen limited, but have never evolved the ability to fix nitrogen, probably because of both phylogenetic and metabolic constraints. Instead, many plants have evolved symbioses with different groups of nitrogen-fixing bacteria (Franche et al. 2009) (Table 1.1). Evolving a symbiosis with bacteria appears to be simpler than evolving a new metabolic function. However, even the evolution of the symbiosis appears constrained within subsets of the larger phylogeny of plants. Even though greater access to fixed nitrogen would probably increase the fitness of most plants, only four orders of rosids associate with nitrogen fixing bacteria (Gherbi et al. 2008). Constraints on the evolution of interactions remain poorly understood (Box 3). Dependence on another organism clearly entails risks (Colwell et al. 2012), but in certain contexts the same kinds of associations emerge repeatedly and independently. Exploring convergent interactions will likely provide new insights into how phylogenetic constraints have shaped and continue to shape the evolution of multispecies associations.

Convergent interactions are unlikely to result from neutral, stochastic processes. Losos (2011) discusses three alternative mechanisms that would result in apparent convergent evolution. Traits might be convergent as a result of coincidence (a spurious correlation due to random chance), or exaptation (a feature that originally evolved in response to a different selective pressure), or because traits are a correlated response to selection on a different character (when similar constraints are shared by taxa, responses can also be shared) (Losos 2011). An evolved interaction between species is unlikely to result from random chance. While an interaction may be interpreted as involving exaptation—e.g. mutualisms have evolved from parasitisms (Weeks et al. 2007) and symbionts have switched hosts (Wolfe et al. 2010)—if the same changes happen repeatedly and independently among different, geographically disparate groups of organisms, then natural selection is likely at play. It is also difficult to imagine

convergent interactions emerging as a result of selection on correlated traits. But however unlikely, whether convergent interactions are ever the result of neutral processes remains an open question (Box 3).

Conclusions

Interactions within different systems can be defined as convergent if the interactions: 1) evolved independently; 2) involve organisms from different trophic levels; and 3) are functionally similar. Convergent interactions are easier to identify when they are found in geographically separated ecosystems, or occur among organisms with convergently evolved morphological or ecological traits (e.g., herbivores with hindguts or insects that grow fungi).

Convergent interactions provide evidence that natural selection can repeatedly favor certain types of interspecific relationships, and in fitness landscapes involving multiple species, key interactions can represent adaptive peaks. Recognizing convergent interactions provides a framework to generate hypotheses about ecological relationships among poorly studied taxa and to identify potential selective pressures structuring the diversity and function of multispecies interactions across kingdoms. Extrapolating from known to unknown might be most straightforward in systems where natural selection has reciprocally shaped all interacting parties, as coevolving partners are likely to be exerting specific and similar selective pressures on each other.

Our examples reflect our own experiences working with microbes, mutualisms, and food webs, but the concept of convergent interactions will also be useful for understanding other types of interactions. For example, convergence appears to be a feature of parasitic fig wasp communities associated with tree species in Africa, Australia and America (Segar et al. 2013)

(Table 1.1). Species richness differs among the different communities, but the proportions of individual insects within various functional groups are similar across the communities, and these three geographically separated, multi-trophic systems fit the definition of convergent interactions.

Convergent interactions provide a useful framework for interpreting recent discoveries of functional convergence patterns, particularly those involving microbes and animals. For example, like the complex bacterial communities associated with mammalian hindguts and fungus-growing insects, functional equivalence and evolutionary convergence have also been found in the bacterial communities of sponges (Fan et al. 2012) (Table 1.1). An explicit focus on convergent interactions will almost certainly illuminate similar functional relationships influencing community assembly in a myriad of other systems.

Table 1.1. In addition to the five examples detailed in the text, other potential examples of convergent interactions include:

Interaction	Description
Rosid plants and nitrogen-fixing bacteria	Plants in four orders of the rosid clade of angiosperms have close, prolonged associations with nitrogen-fixing bacteria in the Alphaproteobacteria, Betaproteobacteria, and Actinobacteria (Franche et al. 2009). The single evolution of a gene necessary for nodulation likely allowed plants to form associations with bacteria (Gherbi et al. 2008), and the different phyla of bacteria likely acquired nitrogenase genes via horizontal transmission (Raymond et al. 2004); however, the interactions of the plants with different groups of bacteria seem to have emerged independently.
Different functional groups of wasps living within figs	Communities of parasitoid, pollinator, and galler wasps from different evolutionary lineages live in figs in Australia, Africa, and South America (Segar et al. 2013).
Sponges with bacterial symbionts	The different microbial communities that associate with divergent groups of sponges are functionally equivalent (Fan et al. 2012).
Plant pollination syndromes	Plants from different lineages have converged on floral traits that attract particular groups of pollinators. A recent quantitative meta-analysis finds strong evidence that pollination syndromes predict the most efficient pollinators, particularly when pollinators are bats, bees, birds, or moths (Rosas-Guerrero et al. 2014).
Plants producing floral oils and oil-collecting bees	Oil-producing flowers evolved at least 28 times within 11 different plant families, and oil-foraging behavior evolved at least 7 times within bees (Renner and Schaefer 2010).
Marine organisms and bioluminescent bacteria	Bioluminescent symbioses with bacteria in the Vibrionaceae family appear to have evolved independently in four teleost fish orders (Dunlap et al. 2007) and in two squid families (Pankey et al. 2014).

Box 1. **Tools and metrics for identifying convergent interactions**

In some cases, different groups of interacting organisms do not share common ancestors, and convergent interactions are obvious. In other cases, convergent interactions are less obvious but can be identified using a combination of ecological and phylogenetic methods.

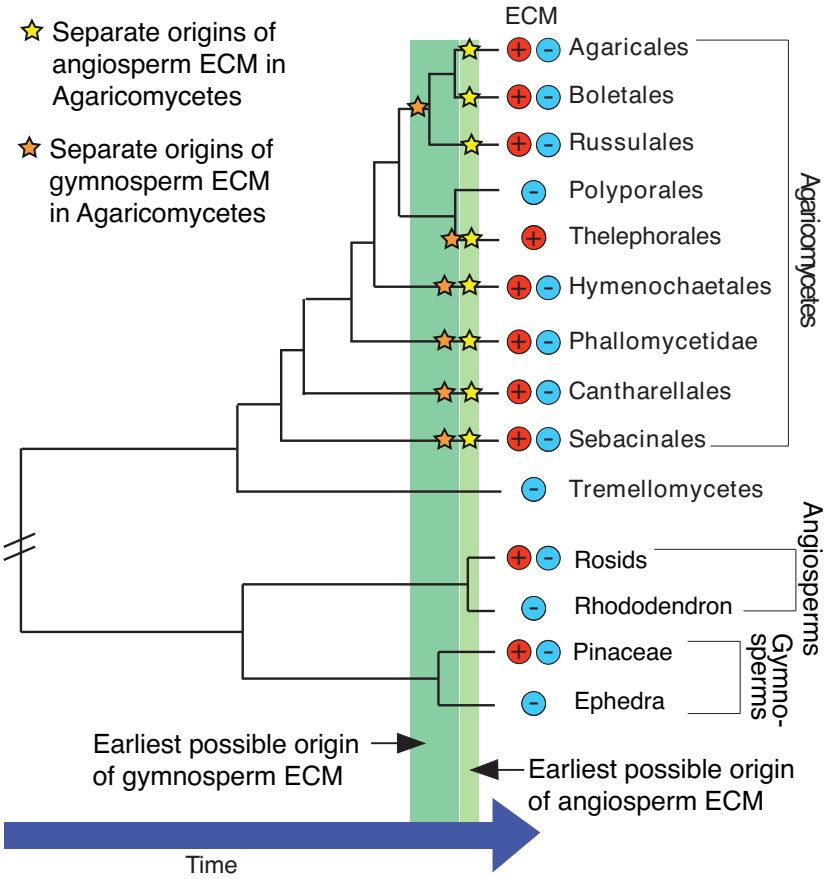
Method 1: Natural History

The simplest way to identify convergent interactions is to describe 1) the interaction, 2) the species involved, and 3) the relationships among species of each interacting group. For example, interactions of ant-plants (from 19 different families) and their ant inhabitants (from 5 different subfamilies) involve the exchange of nesting space and food for protection from herbivores. Phylogenetically independent associations are found on different continents (Figure 1.1). The common ancestor of *Macaranga* and *Cecropia* trees was not a myrmecophyte, and the common ancestor of *Crematogaster* and *Azteca* ants was not an obligate tree-associated ant. Each interaction evolved independently.

Method 2: Phylogenetic Molecular Dating Analysis

Phylogenetic methods are critical for dating the relative ages of clades, and can be used to identify convergent interactions. A necessary postulate is that an interaction cannot evolve before the interacting organisms exist. Certain clades of fungi, for example the Cantharellales, evolved before the appearance of pines or flowering plants in the rosid clade (Hibbett and Matheny 2009) (Box Figure I). Pines and rosids do not share an ectomycorrhizal ancestor, but plants from each clade are ectomycorrhizal. Therefore pine-Cantharellales symbioses must have evolved independently from rosid-Cantharellales symbioses. A phylogeny of agaricomycete fungi and

plants identifies at least eight independent origins of associations between Agaricomycota and angiosperms, and six independent origins between Agricomycota and gymnosperms (Hibbett and Matheny 2009) (Box Figure I). Phylogenetic dating is straightforward when interactions involve two organisms, but more difficult to use in systems involving three or more species.



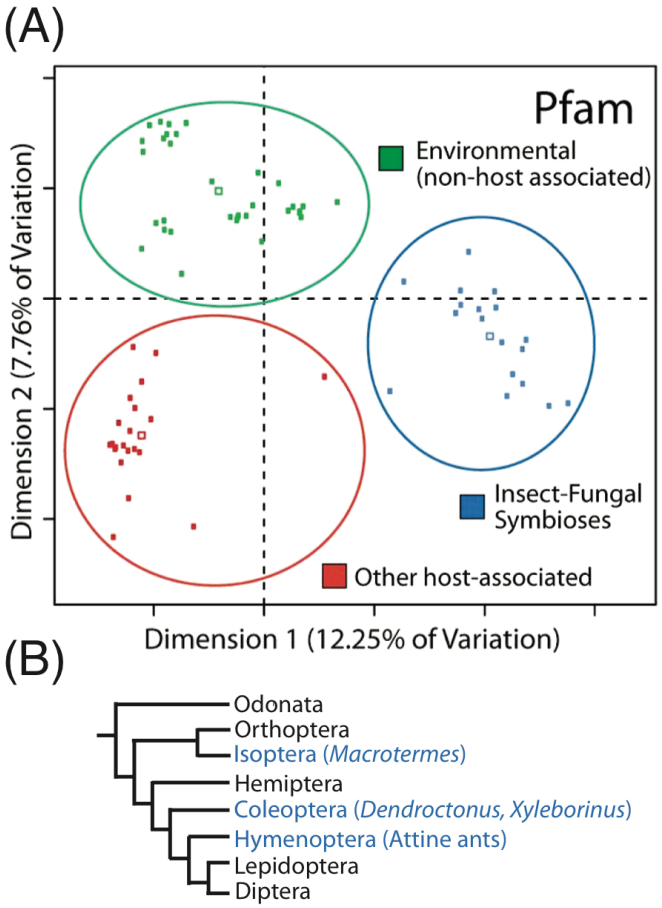
Box 1, Figure I: Multiple independent evolutions of ectomycorrhizal symbioses

Legend: Multiple independent origins of ectomycorrhizal (ECM) associations between agaricomycete fungi and angiosperms or gymnosperms (yellow or orange stars, respectively). Rosids and pines do not share a common ancestor associated with ectomycorrhizal fungi, and as these taxa evolved after certain clades of ectomycorrhizal agaricomycete fungi, the symbiotic interactions must have arisen independently. Modifed from (Hibbett and Matheny 2009) under Creative Commons Attribution (CC-BY) License © 2009 Hibbett and Matheny.

Method 3: Ordination

Ordination can identify convergence among microbial communities and their hosts by clustering communities according to functional similarity (Box Figure II). Ordination methods

require information about 1) host taxonomy; 2) host traits (e.g., gut morphology or fungal cultivation); and 3) the presence, abundance, and functional traits of microbes. Clusters can be identified visually and tested using permutational multivariate analysis of variance (Anderson 2001). When communities associated with hosts from different lineages cluster together according to a convergent host trait, the interactions of the hosts and their communities are recognized as convergent (Aylward et al. 2014) (Box Figure II).



Box 1, Figure II: Functional convergence in bacterial communities of fungus-growing insects

Legend: Metagenomic functional profiles of bacterial communities associated with convergently evolved insect-fungal symbioses cluster separately from environmental or other host associated communities. A) Principal coordinate analysis of bacterial community metagenomes annotated using the Protein Families (Pfam) database. B) Simplified phylogeny of select insect orders. Orders including insects with insect-fungal symbioses shown in (A) are highlighted in blue. Modifed from (Aylward et al. 2014) under Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license © 2014 Aylward et al.

Box 2. Potential methods for exploring patterns of convergent interactions

With sufficient data, recently developed methods for mapping trait evolution onto phylogenies (Ingram and Mahler 2013) or for quantifying the strength of convergent evolution (Arbuckle et al. 2014) might emerge as useful resources for identifying and measuring convergent interactions. However, to use these tools, one set of interacting partners must be redefined as a trait of interest or measure of diversity. For example, continuous functional traits (e.g., cellulose degradation) and bacterial gut community diversity measures (e.g., phylogenetic beta diversity) could be mapped onto a host phylogeny (e.g., mammals) using SURFACE (Ingram and Mahler 2013). SURFACE identifies convergent evolutionary regimes and then uses simulations to test whether there is more convergence than would be expected by chance. Using this approach, one could test whether herbivores have associated convergently with bacterial communities having high cellulose-degradation capacities. Adding the recently developed Wheatsheaf index can enable quantification of the strength of convergence, by measuring phenotypic similarity while penalizing for phylogenetic relatedness (Arbuckle et al. 2014).

Network analyses also can be used to describe convergent patterns of interactions across multispecies communities. To date, analyses exploring convergence across multiple networks have focused on very broad network structure; for example, when looking at different kinds of networks, plant-animal mutualistic networks are more highly nested than food-webs, regardless of the type of mutualism (Bascompte et al. 2003). Within networks, selection on a complementary trait between trophic levels can lead to trait convergence within a trophic level (Guimarães et al. 2011). The connection between coevolution and trait convergence in mutualistic networks is complex, with results depending on the strength of selection (Nuismer et

al. 2013). Similar network approaches likely can be extended to comparisons across multiple networks to identify convergent interactions, where different species in different systems have similar functional roles. For example, two systems with convergent interactions might have network topologies that are more similar than expected by chance alone, if different organisms fulfilling the same ecosystem function exhibit similar measures of centrality.

Advances in metagenomic sequencing and stable isotope analysis (particularly stable isotope probing (Dumont and Murrell 2005)) can provide additional information on functional genes and trophic levels for small, difficult to observe organisms such as microbes and invertebrates (Layman et al. 2012, Haig et al. 2015). New technologies and emerging protocols should allow microbes to be incorporated into existing food webs of plants and animals, which can then be used to explore convergence across entire ecosystem networks.

Box 3: Outstanding Questions

- How common are convergent interactions?
- When do organisms evolve the capacity for a particular function *versus* evolving an interaction with a different organism that can already perform that function?
- How often are convergent interactions evolutionary innovations, and the cause of increased niche breadth and perhaps ultimately adaptive radiations?
- How do the population dynamics of species affect the emergence of convergent interactions? The demographics of individuals among populations may speed or slow the emergence of interactions.
- What constrains the evolution of convergent interactions? Does convergence require particular environments, for example habitats lacking in a key resource, or traits with simple genetic underpinnings? How do organisms' developmental and phylogenetic constraints affect the emergence of convergent interactions?
- Can convergent interactions result solely from neutral processes?

- What can we learn from identifying differences among convergent interactions, for example, differences among ectomycorrhizal symbioses that have evolved independently?

- What new methods are required to identify, understand, and quantify convergent interactions?

Chapter 1 Acknowledgments

Our research was supported by an NSF Graduate Research Predoctoral Fellowship and NSF Doctoral Dissertation Grant DEB-1400982 (to L.S.B.), NSF grants DEB 1144056 and DEB 1237491 (to A.M.E), a grant from the HFSP (to A.P.), and a Templeton FQEB grant (to N.E.P. and A.P). We thank Jon Sanders, Charles Davis, Dan Sternof Beyer, Tom Miller, and Matt Lau for helpful discussions and suggestions. We appreciate the feedback from three anonymous reviewers, which greatly improved the manuscript.

Chapter 1 References:

Adlassnig, W., M. Peroutka, and T. Lendl. 2011. Traps of carnivorous pitcher plants as a habitat: composition of the fluid, biodiversity and mutualistic activities. Annals of botany 107:181–194.

Albert, V. A., S. E. Williams, and M. W. Chase. 1992. Carnivorous plants: phylogeny and structural evolution. Science 257:1491–1495.

An, C.-I., E.-I. Fukusaki, and A. Kobayashi. 2001. Plasma-membrane H+-ATPases are expressed in pitchers of the carnivorous plant *Nepenthes alata* Blanco. Planta 212:547–555.

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. Austral ecology 26:32–46.

Arber, A. 1941. On the morphology of the pitcher-leaves in *Heliamphora*, *Sarracenia*, *Darlingtonia*, *Cephalotus*, and *Nepenthes*. Annals of botany 5:563–578.

Arbuckle, K., C. M. Bennett, and M. P. Speed. 2014. A simple measure of the strength of convergent evolution. Methods in Ecology and Evolution 5:685–693.

Arnold, A. E., L. C. Mejía, D. Kyllo, E. I. Rojas, Z. Maynard, N. Robbins, and E. A. Herre. 2003. Fungal endophytes limit pathogen damage in a tropical tree. Proceedings of the National Academy of Sciences of the United States of America 100:15649–15654.

Aylward, F. O., G. Suen, P. H. W. Biedermann, A. S. Adams, J. J. Scott, S. A. Malfatti, T. Glavina del Rio, S. G. Tringe, M. Poulsen, K. F. Raffa, K. D. Klepzig, and C. R. Currie. 2014. Convergent bacterial microbiotas in the fungal agricultural systems of insects. mBio 5:e02077.

Bares, J. C., D. D. Ackerly, D. A. Baum, and F. A. Bazzaz. 2004. Phylogenetic overdispersion in Floridian oak communities. The American Naturalist 163:823–843.

Barke, J., R. F. Seipke, S. Grüschow, D. Heavens, N. Drou, M. J. Bibb, R. J. M. Goss, D. W. Yu, and M. I. Hutchings. 2010. A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant *Acromyrmex octospinosus*. BMC biology 8:109.

Bascompte, J., P. Jordano, C. J. Melián, and J. M. Olesen. 2003. The nested assembly of plant-animal mutualistic networks. Proceedings of the National Academy of Sciences of the United States of America 100:9383–9387.

Beimforde, C., N. Schäfer, H. Dörfelt, P. C. Nascimbene, H. Singh, J. Heinrichs, J. Reitner, R. S. Rana, and A. R. Schmidt. 2011. Ectomycorrhizas from a Lower Eocene angiosperm forest. New Phytologist 192:988–996.

Bittleston, L. S., F. Brockmann, W. Wcislo, and S. A. Van Bael. 2011. Endophytic fungi reduce leaf-cutting ant damage to seedlings. Biology letters 7:30–32.

Bradshaw, W. E., and R. A. Creelman. 1984. Mutualism between the carnivorous purple pitcher plant and its inhabitants. American Midland Naturalist 112:294–304.

Bronstein, J. L. 2006. The contribution of ant-plant protection studies to our understanding of mutualism. Biotropica 30:150–161.

Butler, J. L., N. J. Gotelli, and A. M. Ellison. 2008. Linking the brown and green: nutrient transformation and fate in the *Sarracenia* microecosystem. Ecology 89:898–904.

Cafaro, M. J., and C. R. Currie. 2005. Phylogenetic analysis of mutualistic filamentous bacteria associated with fungus-growing ants. Canadian Journal of Microbiology 51:441–446.

Cardoza, Y. J., K. D. Klepzig, and K. F. Raffa. 2006. Bacteria in oral secretions of an endophytic insect inhibit antagonistic fungi. Ecological Entomology 31:636–645.

Cavender-Bares, J., K. H. Kozak, P. V. A. Fine, and S. W. Kembel. 2009. The merging of community ecology and phylogenetic biology. Ecology Letters 12:693–715.

Clay, K. 1988. Fungal endophytes of grasses: a defensive mutualism between plants and fungi. Ecology 69:10–16.

Clements, F. E. 1916. Plant Succession: An Analysis of the Development of Vegetation. Carnegie Institution of Washington, Washington D.C.

Clements, F. E. 1936. Nature and structure of the climax. Journal of Ecology 24:252–284.

Cody, M. L., and H. A. Mooney. 1978. Convergence versus nonconvergence in mediterranean-climate ecosystems. Annual Review of Ecology and Systematics 9:265–321.

Colwell, R. K., R. R. Dunn, and N. C. Harris. 2012. Coextinction and persistence of dependent species in a changing world. Annual Review of Ecology, Evolution, and Systematics 43:183–203.

Conant, G. C., and A. Wagner. 2003. Convergent evolution of gene circuits. Nature genetics 34:264–266.

Currie, C. R., M. Poulsen, J. Mendenhall, J. J. Boomsma, and J. Billen. 2006. Coevolved crypts and exocrine glands support mutualistic bacteria in fungus-growing ants. Science 311:81–83.

Darwin, C. 1859. On the origin of species. Murray, London.

Davidson, D. W., and D. McKey. 1993. The evolutionary ecology of symbiotic ant-plant relationships. Journal of Hymenopteran Research 2:13–83.

Delsuc, F., J. L. Metcalf, L. Wegener Parfrey, S. J. Song, A. González, and R. Knight. 2014. Convergence of gut microbiomes in myrmecophagous mammals. Molecular Ecology 23:1301–1317.

Douglas, A. E. 2015. Multiorganismal insects: diversity and function of resident microorganisms. Annual review of entomology 60:17–34.

Dumont, M. G., and J. C. Murrell. 2005. Stable isotope probing - linking microbial identity to function. Nature Reviews Microbiology 3:499–504.

Dunlap, P. V., J. C. Ast, S. Kimura, A. Fukui, T. Yoshino, and H. Endo. 2007. Phylogenetic analysis of host–symbiont specificity and codivergence in bioluminescent symbioses. Cladistics 23:507–532.

Ellison, A. M., N. J. Gotelli, J. S. Brewer, D. L. Cochran-Stafira, J. M. Kneitel, T. E. Miller, A. C. Worley, and R. Zamora. 2003. The evolutionary ecology of carnivorous plants. Advances in Ecological Research 33:1–74.

Fan, L., D. Reynolds, M. Liu, M. Stark, S. Kjelleberg, N. S. Webster, and T. Thomas. 2012. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. Proceedings of the National Academy of Sciences of the United States of America 109:E1878–87.

Fiala, B., U. Maschwitz, T. Y. Pong, and A. J. Helbig. 1989. Studies of a South East Asian ant-plant association: protection of *Macaranga* trees by *Crematogaster borneensis*. Oecologia 79:463–470.

Franche, C., K. Lindström, and C. Elmerich. 2009. Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. Plant and Soil 321:35–59.

Frank, J. H. 1983. Phytotelmata. (J. H. Frank and L. P. Lounibos, Eds.). Plexus Pub.

Fukami, T., T. Martijn Bezemer, S. R. Mortimer, and W. H. Putten. 2005. Species divergence and trait convergence in experimental plant community assembly. Ecology Letters 8:1283–1290.

Funk, D. J., L. Helbling, J. J. Wernegreen, and N. A. Moran. 2000. Intraspecific phylogenetic congruence among multiple symbiont genomes. Proceedings of the Royal Society B: Biological Sciences 267:2517–2521.

Gherbi, H., K. Markmann, S. Svistoonoff, J. Estevan, D. Autran, G. Giczey, F. Auguy, B. Peret, L. Laplaze, C. Franche, M. Parniske, and D. Bogusz. 2008. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and *Frankia* bacteria. Proceedings of the National Academy of Sciences of the United States of America 105:4928–4932.

Gleason, H. A. 1926. The individualistic concept of the plant association. Bulletin of the Torrey Botanical Club 53:7–26.

Godoy-Vitorino, F., K. C. Goldfarb, U. Karaoz, S. Leal, M. A. Garcia-Amado, P. Hugenholtz, S. G. Tringe, E. L. Brodie, and M. G. Dominguez-Bello. 2011. Comparative analyses of foregut and hindgut bacterial communities in hoatzins and cows. The ISME Journal 6:531–541.

Guimarães, P. R., P. Jordano, and J. N. Thompson. 2011. Evolution and coevolution in mutualistic networks. Ecology Letters 14:877–885.

Haig, S.-J., M. Schirmer, R. D'Amore, J. Gibbs, R. L. Davies, G. Collins, and C. Quince. 2015. Stable-isotope probing and metagenomics reveal predation by protozoa drives *E. coli* removal in slow sand filters. The ISME Journal 9:797–808.

Heath, J. J., and J. O. Stireman. 2010. Dissecting the association between a gall midge, *Asteromyia carbonifera*, and its symbiotic fungus, *Botryosphaeria dothidea*. Entomologia Experimentalis et Applicata 137:36–49.

Hibbett, D. S., and P. B. Matheny. 2009. The relative ages of ectomycorrhizal mushrooms and their plant hosts estimated using Bayesian relaxed molecular clock analyses. BMC biology 7:13.

Humboldt, von, A., and A. Bonpland. 1805. Essai sur la géographie des plantes. Chez Levrault, Schoell et Compagnie, Libraries, Paris.

Ingram, T., and D. L. Mahler. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods in Ecology and Evolution 4:416–425.

Janzen, D. H. 1966. Coevolution of mutualism between ants and acacias in Central America. Evolution 20:249–275.

Janzen, D. H. 1969. Allelopathy by myrmecophytes: the ant *Azteca* as an allelopathic agent of *Cecropia*. Ecology 50:147–153.

Kau, A. L., P. P. Ahern, N. W. Griffin, A. L. Goodman, and J. I. Gordon. 2011. Human nutrition, the gut microbiome and the immune system. Nature 474:327–336.

Kitching, R. L. 2000. Food webs and container habitats. Cambridge University Press.

Layman, C. A., M. S. Araujo, R. Boucek, C. M. Hammerschlag-Peyer, E. Harrison, Z. R. Jud, P. Matich, A. E. Rosenblatt, J. J. Vaudo, L. A. Yeager, D. M. Post, and S. Bearhop. 2012. Applying stable isotopes to examine food-web structure: an overview of analytical tools. Biological Reviews 87:545–562.

LePage, B. A., R. S. Currah, R. A. Stockey, and G. W. Rothwell. 1997. Fossil ectomycorrhizae from the middle Eocene. American Journal of Botany 84:410–412.

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

Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, and R. Knight. 2008b. Evolution of mammals and their gut microbes. Science 320:1647–1651.

Losos, J. B. 2011. Convergence, adaptation, and constraint. Evolution 65:1827–1840.

Mayer, V. E., M. E. Frederickson, D. McKey, and R. Blatrix. 2014. Current issues in the evolutionary ecology of ant-plant symbioses. New Phytologist 202:749–764.

McFall-Ngai, M., M. G. Hadfield, T. C. G. Bosch, H. V. Carey, T. Domazet-Lošo, A. E. Douglas, N. Dubilier, G. Eberl, T. Fukami, S. F. Gilbert, U. Hentschel, N. King, S. Kjelleberg, A. H. Knoll, N. Kremer, S. K. Mazmanian, J. L. Metcalf, K. Nealson, N. E. Pierce, J. F. Rawls, A. Reid, E. G. Ruby, M. Rumpho, J. G. Sanders, D. Tautz, and J. J. Wernegreen. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proceedings of the National Academy of Sciences of the United States of America 110:3229–3236.

Mooney, H. A., and E. L. Dunn. 1970. Convergent evolution of Mediterranean-climate evergreen sclerophyll shrubs. Evolution 24:292–303.

Moreau, C. S., C. D. Bell, R. Vila, S. B. Archibald, and N. E. Pierce. 2006. Phylogeny of the ants: diversification in the age of angiosperms. Science 312:101–104.

Muegge, B. D., J. Kuczynski, D. Knights, J. C. Clemente, A. González, L. Fontana, B. Henrissat, R. Knight, and J. I. Gordon. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science 332:970–974.

Mueller, U. G., T. R. Schultz, C. R. Currie, R. M. Adams, and D. Malloch. 2001. The origin of the attine ant-fungus mutualism. The Quarterly review of biology 76:169–197.

Nuismer, S. L., P. Jordano, and J. Bascompte. 2013. Coevolution and the Architecture of Mutualistic Networks. Evolution 67:338–354.

Pankey, M. S., V. N. Minin, G. C. Imholte, M. A. Suchard, and T. H. Oakley. 2014. Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. Proceedings of the National Academy of Sciences of the United States of America 111:E4736–42.

Peay, K. G., P. G. Kennedy, S. J. Davies, S. Tan, and T. D. Bruns. 2009. Potential link between plant and fungal distributions in a dipterocarp rainforest: community and phylogenetic structure of tropical ectomycorrhizal fungi across a plant and soil ecotone. New Phytologist 185:529–542.

Quek, S.-P., S. J. Davies, P. S. Ashton, T. Itino, and N. E. Pierce. 2007. The geography of diversification in mutualistic ants: a gene's-eye view into the Neogene history of Sundaland rain forests. Molecular Ecology 16:2045–2062.

Raymond, J., J. L. Siefert, C. R. Staples, and R. E. Blankenship. 2004. The natural history of nitrogen fixation. Molecular biology and evolution 21:541–554.

Renner, S. S., and H. Schaefer. 2010. The evolution and loss of oil-offering flowers: new insights from dated phylogenies for angiosperms and bees. Philosophical Transactions of the Royal

Society B: Biological Sciences 365:423–435.

Rico-Gray, V., and P. S. Oliveira. 2007. The ecology and evolution of ant-plant interactions. University of Chicago Press.

Rohfritsch, O. 2008. Plants, gall midges, and fungi: a three-component system. Entomologia Experimentalis et Applicata 128:208–216.

Rosas-Guerrero, V., R. Aguilar, S. Martén-Rodríguez, L. Ashworth, M. Lopezaraiza-Mikel, J. M. Bastida, and M. Quesada. 2014. A quantitative review of pollination syndromes: do floral traits predict effective pollinators? Ecology Letters 17:388–400.

Rosumek, F. B., F. A. O. Silveira, F. S Neves, N. P. U Barbosa, L. Diniz, Y. Oki, F. Pezzini, G. W. Fernandes, and T. Cornelissen. 2009. Ants on plants: a meta-analysis of the role of ants as plant biotic defenses. Oecologia 160:537–549.

Samuels, C. L., and J. A. Drake. 1997. Divergent perspectives on community convergence. Trends in ecology & evolution 12:427–432.

Schluter, D., and R. E. Ricklefs. 1993. Convergence and the regional component of species diversity. Pages 230–240 *in* R. E. Ricklefs and D. Schluter, editors. Species diversity: historical and geographical perspectives. University of Chicago Press, Chicago.

Scott, J. J., D.-C. Oh, M. C. Yuceer, K. D. Klepzig, J. Clardy, and C. R. Currie. 2008. Bacterial protection of beetle-fungus mutualism. Science 322:63.

Segar, S. T., R. A. S. Pereira, S. G. Compton, and J. M. Cook. 2013. Convergent structure of multitrophic communities over three continents. Ecology Letters 16:1436–1445.

Smith, S. E., and D. J. Read. 2010. Mycorrhizal Symbiosis. Third Edition. Academic Press.

Srivastava, D. S., J. Kolasa, J. Bengtsson, A. Gonzalez, S. P. Lawler, T. E. Miller, P. Munguia, T. Romanuk, D. C. Schneider, and M. K. Trzcinski. 2004. Are natural microcosms useful model systems for ecology? Trends in ecology & evolution 19:379–384.

Stayton, C. T. 2008. Is convergence surprising? An examination of the frequency of convergence in simulated datasets. Journal of theoretical biology 252:1–14.

Stevens, C. E., and I. D. Hume. 1995. Comparative physiology of the vertebrate digestive system. Cambridge University Press.

Thompson, J. N. 1999. The Evolution of Species Interactions. Science 284:2116–2118.

Thompson, J. N. 2005. The geographic mosaic of coevolution. University of Chicago Press.

Um, S., A. Fraimout, P. Sapountzis, D.-C. Oh, and M. Poulsen. 2013. The fungus-growing termite *Macrotermes natalensis* harbors bacillaene-producing *Bacillus sp.* that inhibit potentially antagonistic fungi. Scientific reports 3:3250.

Walter, J., R. A. Britton, and S. Roos. 2011. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. Proceedings of the National Academy of Sciences of the United States of America 108 Suppl 1:4645–4652.

Weeks, A. R., M. Turelli, W. R. Harcombe, K. T. Reynolds, and A. A. Hoffmann. 2007. From parasite to mutualist: Rapid evolution of *Wolbachia* in natural populations of *Drosophila*. PLoS biology 5:997–1005.

Wilson, E. O., and B. Hölldobler. 2005. The rise of the ants: a phylogenetic and ecological explanation. Proceedings of the National Academy of Sciences of the United States of America 102:7411–7414.

Wolfe, B. E., F. Richard, H. B. Cross, and A. Pringle. 2010. Distribution and abundance of the introduced ectomycorrhizal fungus *Amanita phalloides* in North America. New Phytologist 185:803–816.

Young, T. P., C. H. Stubblefield, and L. A. Isbell. 1997. Ants on swollen-thorn acacias: species coexistence in a simple system. Oecologia 109:98–107.

Chapter 2: Metabarcoding as a tool for investigating arthropod diversity in *Nepenthes* pitcher plants

Published in Austral Ecology online July 14, 2015. DOI: 10.1111/aec.12271

Authors: Leonora S. Bittleston[1], Christopher C. M. Baker[1], Lila B. Strominger[1], Anne Pringle[2] and Naomi E. Pierce[1]

Addresses:

[1]Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA, 02138 USA

[2]Harvard Forest, Harvard University, 324 North Main Street, Petersham, MA 01366 USA

Corresponding author: L. S. Bittleston

Address: 26 Oxford Street, Cambridge, MA 02138 USA

Emails: lbittles@fas.harvard.edu

Abstract:

The biodiversity of tropical forests consists primarily of small organisms that are difficult to detect and characterise. Next-generation sequencing (NGS) methods can facilitate analyses of these arthropod and microbial communities, leading to a better understanding of existing diversity and factors influencing community assembly. The pitchers of carnivorous pitcher plants often house surprisingly discrete communities, and provide ideal systems for analysis using a NGS approach. The plants digest insects in order to access essential nutrients while growing in poor soils; however, the pitchers are also home to communities of living organisms, called inquilines. Certain arthropods appear to have coevolved with their pitcher plant hosts, and are not found in other environments. We used Illumina amplicon sequencing of 18S rDNA to characterise the eukaryotes in three species of *Nepenthes* (Nepenthaceae) pitcher plants: *N. gracilis, N. rafflesiana,* and *N. ampullaria*, in each of three different parks in Singapore. The data reveal an unexpected diversity of eukaryotes, significant differences in community diversity among host species, variation in host specificity of inquilines, and the presence of gregarine parasites. Counts of whole inquiline arthropods from the first collection year were roughly correlated with scaled 18S sequence abundances, indicating that amplicon sequencing is an effective means of gauging community structure. We barcoded a subset of the dipteran larvae using COI primers, and the resulting phylogenetic tree is mostly congruent with that found using the 18S locus, with the exception of one of five morphospecies. For many 18S and COI sequences, the best BLASTn matches showed low sequence identity, illustrating the need for better databases of Southeast Asian dipterans. Finally, networks of core arthropods and their host species were used to investigate degree of host specificity across multiple hosts, and this revealed significant specialisation of certain arthropod fauna.

Keywords: Carnivorous plant, insect, 18S amplicon sequencing, microcosm, network

Introduction:

Tropical rainforests house an astounding diversity of organisms. Arthropods and microscopic organisms represent the majority of this diversity; however, due to their small sizes, it is difficult and often impractical to characterise their communities using traditional survey methods. Recent efforts to describe the full diversity of arthropods in a Panamanian tropical rainforest found 6,144 species in less than one-half hectare, and estimated that 25,000 arthropod species exist within a 6,000 hectare reserve (Basset *et al.* 2012). The rainforests of Southeast Asia are currently threatened by anthropogenic activities, including the highest relative rates of deforestation compared to other tropical regions (Sodhi *et al.* 2004). Many organisms may lose their habitat before their existence is even recognized, as the vast biodiversity of arthropods and other small eukaryotes in Southeast Asian rainforests is still virtually unknown. Next-generation sequencing (NGS) methods have the potential to reveal a large extent of the total diversity within these rich ecosystems (Hajibabaei et al. 2011, Taberlet et al. 2012).

Carnivorous plants have been recognized as hosts for insects since the 1800's (Riley 1874) and are now model systems for food web and microcosm studies (Kitching 2000, Kneitel and Miller 2002, Srivastava et al. 2004). Inside every pitcher is a small ecosystem, presenting an ideal opportunity for studying contained, clearly defined communities. The modified leaves of pitcher plants form cup-shaped vessels that hold a mix of rainwater and excreted digestive enzymes. Pitcher plants tend to grow in low-nutrient soils, and absorb nitrogen, phosphorus, and potentially other nutrients from digested prey (Chapin and Pastor 1995). Pitchers actively attract

insects with extra-floral nectar and possibly UV reflectance (Moran *et al.* 1999), and trap prey with their slippery inner walls, downward pointing hairs, and pitcher fluid (Adlassnig *et al.* 2011). Although pitcher plants trap and drown prey, they also host populations of aquatic arthropods, protists, bacteria, and fungi, often called 'inquilines' (Kitching 2000). Certain species appear to exist only in pitcher plant habitats and have likely adapted to the conditions in the pitcher (Beaver 1985). The most prominent arthropods living within pitchers are mites and dipteran larvae, and the most common prey items are ants (Kitching 2000, Ellison and Gotelli 2009).

There are three families of pitcher plants: Nepenthaceae, Sarraceniaceae, and Cephalotaceae. The plants have evolved independently from three distinct lineages in three different parts of the world: Southeast Asia, the Americas, and Australia (Albert *et al.* 1992). The family Nepenthaceae has one genus, *Nepenthes*, with over 100 species recognized by the IUCN Red List of Threatened Species (although several are listed as lower risk or least concern). New *Nepenthes* species are still frequently discovered and described (Gronemeyer *et al.* 2014).

Nepenthes pitchers associate with a diversity of organisms, though dipteran insects are the most common macrofauna in the internal food webs (Kitching 2000). The arthropod food webs of *Nepenthes* pitchers vary with geography, and are more complex and species-rich closer to the centre of the genus's distribution (Beaver 1985). The inquilines vary with host species (Clarke and Kitching 1993) and have complex predator-prey dynamics (Mogi and Yong 1992). Some species of dipteran insects and aquatic mites appear to be specialised to *Nepenthes* habitats (Ratsirarson and Silander 1996, Fashing 2002, Fashing and Chua 2002).

Next-generation amplicon sequencing, most commonly used with 16S ribosomal primers to identify the composition of prokaryotic communities, has greatly increased our ability to

characterise microscopic organisms (Caporaso *et al.* 2011). For eukaryotes, in particular soil and marine protists, 18S (the homologue of prokaryotic 16S) rRNA primers are used to elucidate microscopic diversity (Stoeck et al. 2010, Bik et al. 2012). However, amplicon sequencing is not a perfect solution for characterising community structure. Sequences are typically shorter than Sanger sequenced barcodes, and thus contain less taxonomic information. PCR biases can affect final sequence abundances, so that they do not accurately represent the number of organisms in a sample (Acinas *et al.* 2005). Additionally, genomes can contain multiple copies of ribosomal genes; studies estimate prokaryotic genomes have from 1 to 15 16S rRNA gene copies (Klappenbach *et al.* 2001), while eukaryotes can have hundreds or even thousands of copies of 18S rRNA genes, as 18S rRNA copy number scales with genome size (Prokopowich *et al.* 2003). Appropriate methods minimize PCR biases; however, 18S sequence abundances still have to be treated with caution, as robust strategies for dealing with 18S copy number variation have not yet been established. Next-generation sequencing (NGS) methods are now being developed for biodiversity monitoring of larger organisms using the same approach as microbial amplicon sequencing and often called "metabarcoding" in this context (Taberlet et al. 2012, Yu et al. 2012). In this paper, we use the terms "amplicon sequencing" and "metabarcoding" interchangeably. For metabarcoding studies of arthropods the COI gene is often used, as it has the advantage of being a single-copy gene with better taxonomic resolution than 18S (Yu *et al.* 2012). However, the COI gene also has limitations, such as poorly conserved primer binding sites and thus less taxonomic coverage compared to rRNA genes (Deagle *et al.* 2014).

To the best of our knowledge, this study is the first attempt to characterise the complete eukaryotic communities within *Nepenthes* pitcher plants using NGS. We address four main questions in this study. First, is metabarcoding with 18S primers an effective tool for the

characterisation of eukaryotic communities, and specifically arthropods? Second, how does 18S metabarcoding compare to COI Sanger sequencing in terms of taxonomic resolution? Third, are numbers of 18S sequences roughly representative of the number of individuals present in the communities? And finally, can 18S metabarcoding be used to study ecological dynamics and host specialisation in natural communities?

Methods:

Sample collection

Samples of pitcher plant fluid were collected in January 2012 and March 2013 from three different parks in Singapore: Bukit Timah Nature Preserve (BTNP), Kent Ridge Park (KRP), and between Upper and Lower Peirce Reservoir Park (UPR) (Table 2.1). Pitcher fluid was collected from the three species of *Nepenthes* with natural distributions in Singapore: *N. gracilis*, *N. rafflesiana* and *N. ampullaria* (Figure 2.1B). In each park, sites were chosen where all three species coexist within a small region, each about 5 – 30 m across. The UPR site had very few *N. ampullaria*, and no samples were collected for this species in 2012. Pitcher fluid and inquilines were collected using a sterile transfer pipette for each sample, and were stored in sterile tubes. Some of the *N. gracilis* samples had very low volume, and fluid from multiple pitchers on the same plant was pooled (see Supplementary Table 2.1). For the 2013 samples, we recorded the total volume within a pitcher, and removed a small amount of fluid to measure the pH using colorpHast pH strips. We added a cetyl trimethylammonium bromide and salt solution (hereafter 'CTAB'; final concentrations: 2% CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris pH 8) to each sample in the same volume as the sample, in order to preserve DNA. Samples were transported to Harvard University at room temperature in CTAB, and then frozen until

processed. Before extracting DNA from *Nepenthes* pitcher fluid, we used sterilized gauze to separate larger arthropod larvae and prey from the fluid.

Table 2.1. Successful samples from 18S amplicon sequencing across three host species and three collecting sites (see text for explanation)

2012 Samples

	BTNP	KRP	UPR	All sites
N. ampullaria	5	3	0	8
N. gracilis	2	7	1	10
N. rafflesiana	2	6	2	10
All species	9	17	3	28

2013 Samples

	BTNP	KRP	UPR	All sites
N. ampullaria	8	5	3	15
N. gracilis	7	9	4	20
N. rafflesiana	7	6	8	22
All species	22	20	15	57

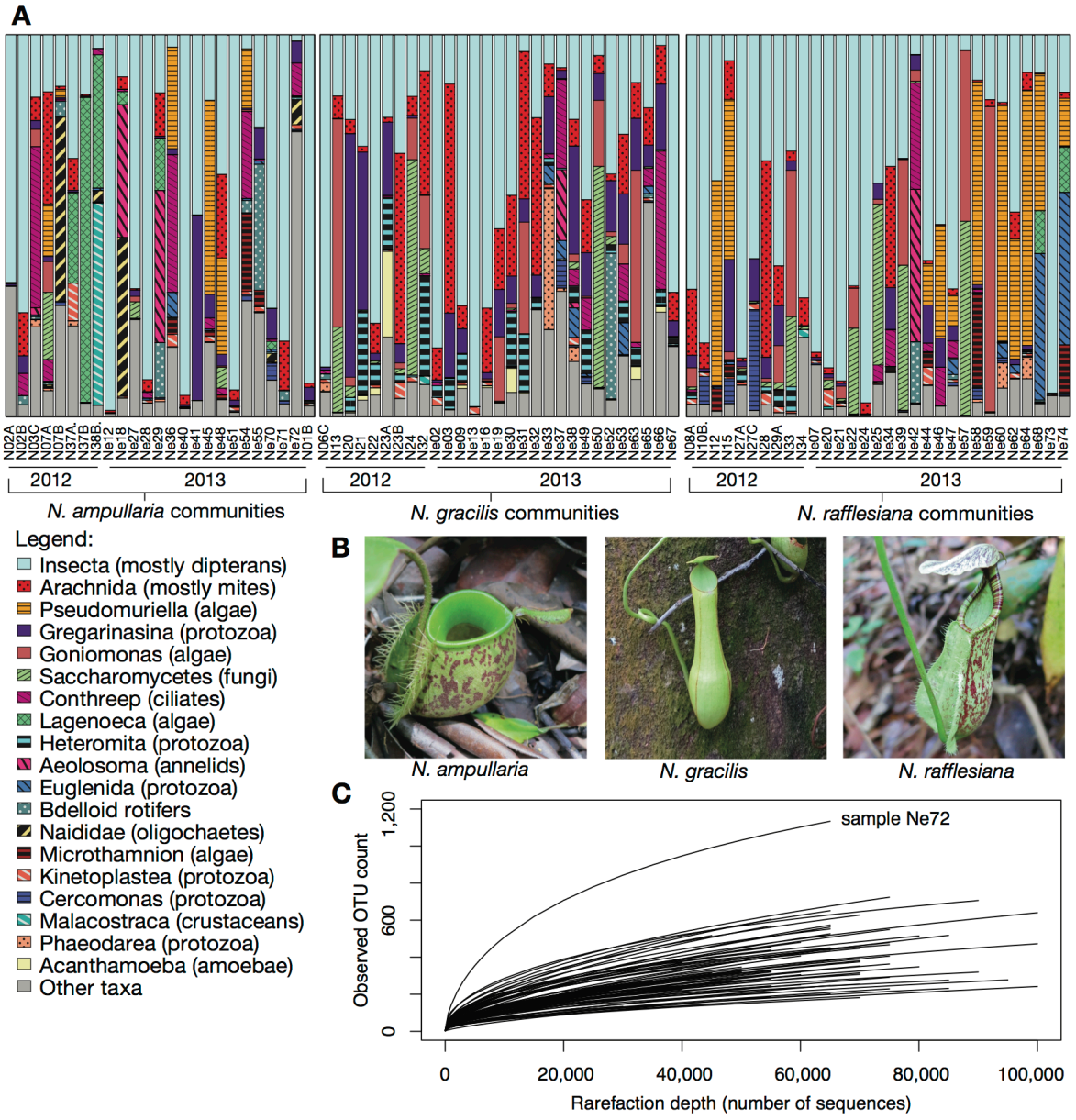


Figure 2.1:
Eukaryotes in *Nepenthes* pitcher plants. A. Summary of the eukaryotic taxa found within the pitcher fluid of three *Nepenthes* species. Each column is one sample, and the y-axis is the relative sequence abundance of the taxa listed in the legend. Samples are grouped by *Nepenthes* species, shown in the photographs (B). C. Rarefaction curves for each sample showing the number of observed OTUs (y-axis) at different sampling depths (x-axis). Sample Ne72 contained soil, and was removed from subsequent analyses.

DNA extraction and 18S amplicon NGS

To concentrate cells, half of the fluid from each sample was either filtered through sterilized 0.22 micron Durapore filters in Swinnex holders (2012 samples) or centrifuged (2013 samples). We extracted DNA from ¾ of each filter or from centrifuged pellets using a phenol-chloroform bead-beating extraction method. These two techniques produced similar results when tested by using both methods on a single sample, and the different approaches are unlikely to affect community diversity analyses (L.S. Bittleston, unpublished data 2013). For each set of DNA extractions we used a negative control to test for contamination. DNA quality was initially evaluated with a Nanodrop spectrophotometer. As some samples had high levels of polyphenols, we cleaned the 2013 samples with a MoBio Powerclean kit. DNA from successfully extracted samples was quantified with a Qubit fluorometer and was sent to Argonne National Laboratories for Illumina MiSeq next-generation amplicon sequencing. The Earth Microbiome Project's barcoded 18S primers were used to amplify eukaryotic DNA (Amaral-Zettler et al. 2009, Caporaso et al. 2012). PCR and sequencing was done according to the Earth Microbiome Project protocols (http://www.earthmicrobiome.org/emp-standard-protocols/18s).

18S quality control and OTU picking

The MiSeq Illumina sequencing output was processed using QIIME 1.8 (Caporaso *et al.* 2010). We split libraries at a quality cut-off of 20, which translates to a base call error rate of 0.01, and then identified and removed chimeras with USEARCH61 (Edgar 2010). After sequencing and quality control, 85 samples were available for analysis (Table 2.1, Supplementary Table 2.1). DNA sequences, averaging 151 nucleotides in length, were clustered into OTUs (Operational Taxonomic Units; here used as a proxy for species) at 97% identity with

reverse strand matching using UCLUST open-reference clustering and the SILVA database for eukaryotes (Pruesse *et al.* 2007). We first assigned taxonomy with the Ribosomal Database Project (RDP) classifier; however, over 40% of our sequences were unassigned, so we then assigned taxonomy with BLAST (Altschul *et al.* 1990). As some sequences were assigned to Bacteria, we split the OTU table at the domain level and continued analyses with only OTUs assigned to Eukaryota. We generated taxa summaries (Figure 2.1A), a rarefaction curve (Figure 2.1C), OTU tables, and initial diversity analyses for all Eukaryota in QIIME, and then filtered and collected OTUs assigned as Arthropoda into an arthropod OTU table in order to analyse these taxa separately. The arthropod OTU table was imported into R in the biom format, and each OTU was given an alphanumeric name according to taxonomy and abundance using an R script we wrote to assign these identities. To avoid over-representation of certain samples, the arthropod OTU table was randomly subsampled to the level of the sample with the fewest sequences: 1,595. We then took the square root of all observations, to decrease the impact of certain OTUs having falsely high abundances due to PCR replication or 18S copy number variation (Prokopowich et al. 2003, Acinas et al. 2005). The square roots of the OTU sequence numbers were used for all downstream analyses.

COI barcoding, and phylogenetic trees of COI and 18S

In order to build a COI phylogenetic tree, dipteran insect larvae and mites from our *Nepenthes* samples were selected for COI barcoding. Individuals were selected to represent a diversity of morphospecies. We extracted DNA from individuals with the Autogen DNA extraction kit and AutoGen Prep 965, amplified COI using LCO1490 and HCO2198 primers (Folmer *et al.* 1994) and the same PCR conditions as a previous study (Hebert *et al.* 2003). We

purified the PCR reaction with AMPure beads (Agencourt) and sequenced with Sanger sequencing. The sequences were quality checked using the program 4Peaks, and good quality sequences were exported as fasta files. All sequences from mites were low quality, likely due to the small size of individuals, so only barcodes from the dipteran insects were used for analyses. Multiples of identical COI sequences were removed. We first assigned taxonomy using the COI Barcode of Life Database (BOLD); however, many of our sequences had no matches, or matched to unnamed sequences. Consequently, sequences were assigned taxonomy using BLASTn and the NCBI database (Supplementary Table 2.3), and aligned with the program MUSCLE (Edgar 2004) via Mesquite (http://mesquiteproject.org). Maximum Likelihood (ML) analysis was performed using the GTRCAT model of evolution and bootstrap resampling (100 replicates) in RAxML-HPC2 version 8.0.24 (Stamatakis 2014) via the CIPRES portal (http://www.phylo.org/sub_sections/portal). The best-scoring ML phylogenetic tree was visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree) (Figure 2.2).

We built an 18S phylogenetic tree to compare with the COI tree and to select OTUs for comparison with morphospecies counts. From the subsampled arthropod table, we used OTUs assigned to dipteran insects with at least 50 sequences per OTU (corresponding to the 24 most abundant dipterans). The representative sequence for each OTU was added to a fasta file. In order to have comparable taxonomy assignments for the OTUs and the COI sequences, each sequence was individually BLASTed to the NCBI nucleotide database (Supplementary Table 2.3). Sequences were aligned and a tree was built as described above for the COI sequences (Figure 2.2).

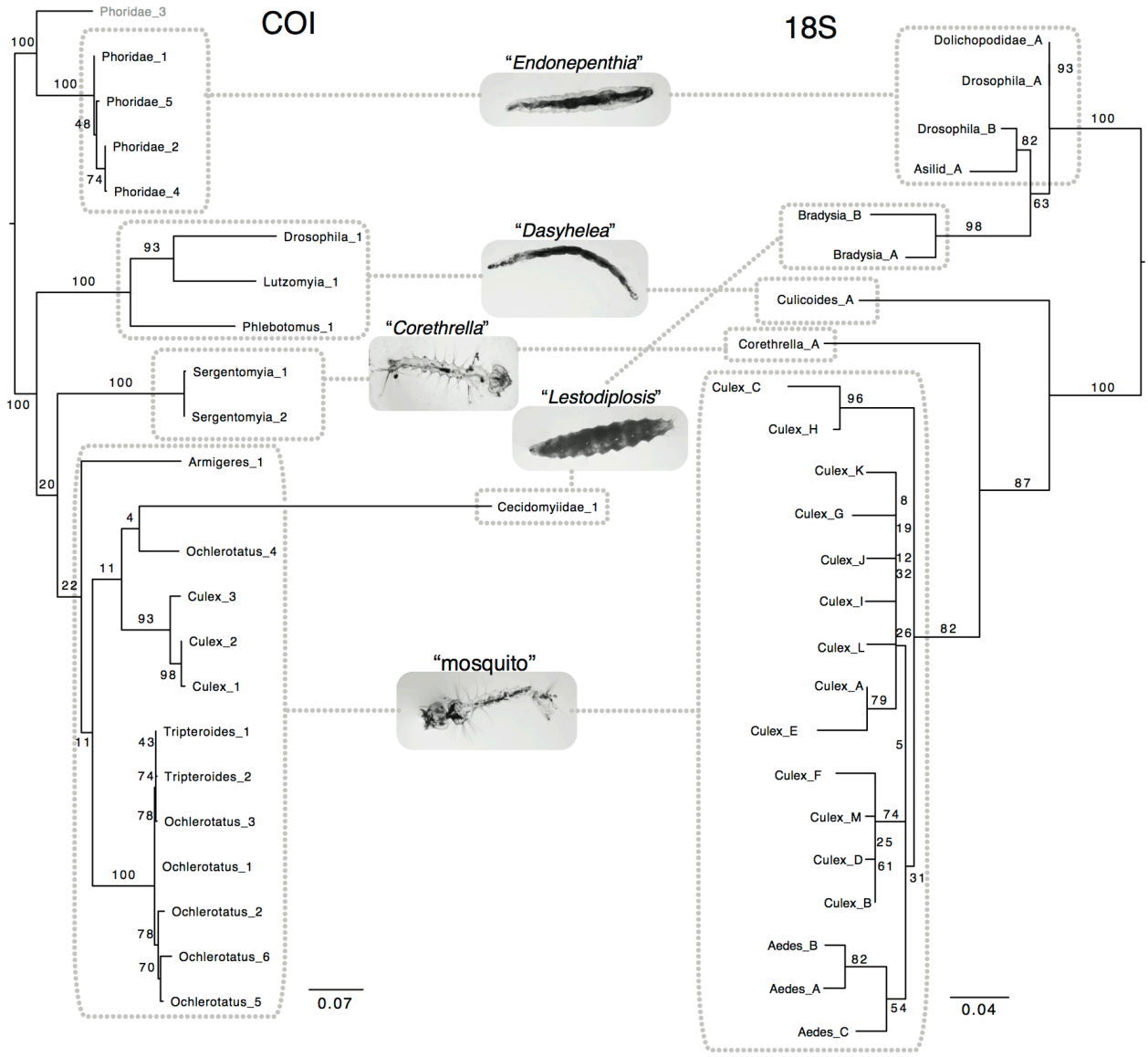


Figure 2.2:
COI and 18S Maximum Likelihood phylogenetic trees with bootstrap values at each node. The clades corresponding to each dipteran morphospecies are outlined and connected across the phylogenies. Taxa on the branch tips are named according to best BLASTn matches.

Arthropod counts and comparison with 18S sequences

Arthropods from a subset of the 2012 samples (Table 2.2) were counted under a dissecting microscope, and we assigned general morphospecies names based on morphological appearance using *A guide to the carnivorous plants of Singapore* (Kai Lok *et al.* 1997). We also took photographs of inquilines under a dissecting microscope for future reference.

Table 2.2. Inquiline individual and sequence counts

	"Corethrella"		"Dasyhelea"		"Endonepenthia"		"Lestodiplosis"		"Mite"		"Mosquito"	
Sample	**Count**	**Seq.†**	**Count**	**Seq.†**	**Count**	**Seq.†**	**Count**	**Seq.†**	**Count**	**Seq.†**	**Count**	**Seq.†**
N01B	0	0	0	0	0	0	0	1.0	0	4.6	3	39.0
N02A	0	0	2	5.7	3	0	0	20.5	0	2.6	5	32.2
N02B	0	0	0	0	7	1.0	0	0	0	13.9	4	36.8
N03C	1	19.8	24	17.0	0	0	0	1.0	0	21.1	1	21.0
N06C	0	0	2	1.7	0	0	0	0	0	0.0	4	39.9
N07A	0	0	1	0	5	12.0	0	0	6	32.5	5	19.7
N07B	0	0	11	14.6	0	0	0	0	12	6.6	2	18.9
N08A	0	0	0	0	3	3.3	0	0	7	18.7	24	29.5
N10B.	0	0	0	0	0	0	0	0	0	10.8	14	35.8
N12	0	0	0	0	0	0	0	1.0	0	1.4	20	39.3
N20	0	0	0	0	0	0	0	0	15	15.7	2	36.2
N21	0	0	0	0	1	0	4	5.4	3	8.8	10	38.4
N22	0	0	0	0	2	2.2	1	0	1	11.4	16	37.1
N23A	0	0	0	0	0	0	3	1.0	0	8.7	6	36.5
N23B	0	0	0	0	2	6.5	0	0	3	34.0	4	9.8
N24	0	0	0	1.0	3	3.0	0	0	1	19.7	1	28.8
N27A	0	0	0	0	31	19.0	0	0	0	3.9	12	34.6
N27C	0	0	27	6.2	0	0	0	0	0	0	38	39.3
N28	0	0	0	0	0	6.8	0	0	22	33.9	3	9.1
N29A	0	0	0	0	1	34.1	1	1.0	3	16.5	0	4.6
N37A.	0	0	52	29.3	0	1.0	0	0	1	17.4	0	9.7
N37B	0	0	0	5.0	0	0	0	0	2	5.6	6	32.2
N38B.	0	0	0	0	0	0	0	0	4	1.0	14	8.2

† Sequence counts are the square root of the number of sequences from the rarefied OTU table

Using taxonomic assignments and clades from the 18S phylogenetic tree, we selected OTUs representing the counted and COI barcoded morphospecies. Mites were added into the dataset by selecting OTUs from the subsampled arthropod table with over 50 sequences (corresponding to the 20 most abundant mite OTUs). Counts of individual arthropod morphospecies were compared to the number of 18S sequences in the OTUs in each sample (Table 2.2). As noted above, we used the square root of 18S sequence counts from the subsampled OTU table, for the comparison. To generate correlations and regressions, we used linear models in R, and permutational linear models (function *lmp* in the lmPerm package) when the assumption of normality was not met. In order to visualize all the results together, we used a $\log_{10}$-$\log_{10}$ plot (Figure 2.3A). To see if rank-abundance of counts corresponds to rank-abundance of sequences, we plotted the same dataset using the *rankabundance* function in the BiodiversityR package in R (Figure 2.3B).

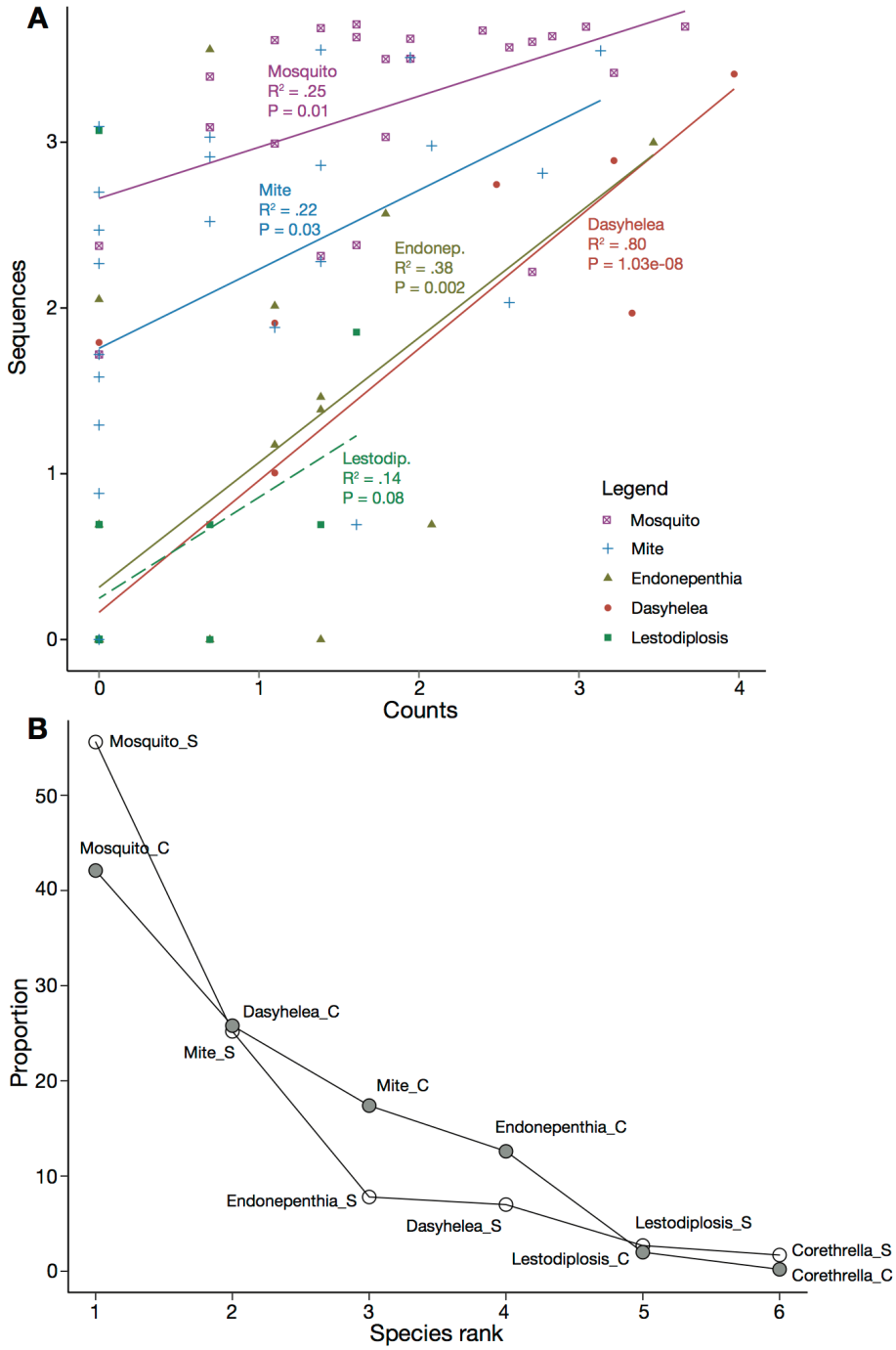


Figure 2.3:
Comparing individuals counts with 18S sequences. A. Scatter plot of inquiline individual counts and 18S sequences plotted on a $\log_{10}$-$\log_{10}$ scale. Regression lines and P-values from the permutational linear models overlie the scatter plot. Solid lines are significant. B. Rank-abundance plot with proportion of community on the y-axis and species rank on the x-axis. An "S" after the morphospecies name and open circles denote 18S sequences, while counts are labeled with a "C" and gray circles.

Multivariate analyses

We tested for significant differences in the 18S OTUs of the arthropod communities by host species using the vegan package in R (Oksanen et al. 2013, R Core Team 2014). Tests were conducted with both permutational multivariate analysis of variance (Anderson 2001) PERMANOVA, *adonis* function) and analysis of similarities (Clarke 1993) ANOSIM, *anosim* function) using Bray-Curtis dissimilarity and the square root of the subsampled OTU counts, as detailed above. These methods can show significant differences when there is different within-group variation (dispersion), so we tested multivariate homogeneity of group dispersions (*betadisper* function) to make sure dispersions were not different among host species groups (Anderson 2006). We also tested whether our pooling of low-volume *N. gracilis* samples had a significant effect on diversity.

Network analyses

We built a bipartite network in order to evaluate the level of specialisation of arthropod inquilines to host species, using the dipteran and mite 18S OTUs from the phylogeny and regression analyses. For the network, we used the presence or absence of subsampled OTUs in each *Nepenthes* host species rather than counts in order not to bias the network with organisms that are very abundant in only one or two samples. Additionally, we removed observations of 3 or fewer sequences, as presence-absence analysis gives equal weight to observations of 1 or 100 and we wanted to avoid skewing the network's specialisation level with very low-abundance observations. The network was built and graphed using the bipartite package in R (Dormann *et al.* 2009) (Figure 2.4). Network-level specialisation was calculated with H_2', a version of the two-dimensional Shannon diversity of the interactions, H_2 (Blüthgen et al. 2006, Dormann

2011). H_2' ranges from 0 (no specialisation) to 1 (all organisms completely specialised). We used a null model based on the Patefield algorithm to test if the network is more specialised than expected under a null distribution, where host species is randomized (Dormann 2011). We checked the robustness of our results by repeating the analyses with counts instead of presence-absence, and with presence-absence without removing observations of 3 and fewer. To further examine the full ecological network, we also analysed a bipartite network including all arthropod OTUs, not solely the inquilines.

A different approach, called a spring-embedded network, was used to investigate core OTUs shared among all host species and many samples (see Supplementary Figure 2.1 and associated methods).

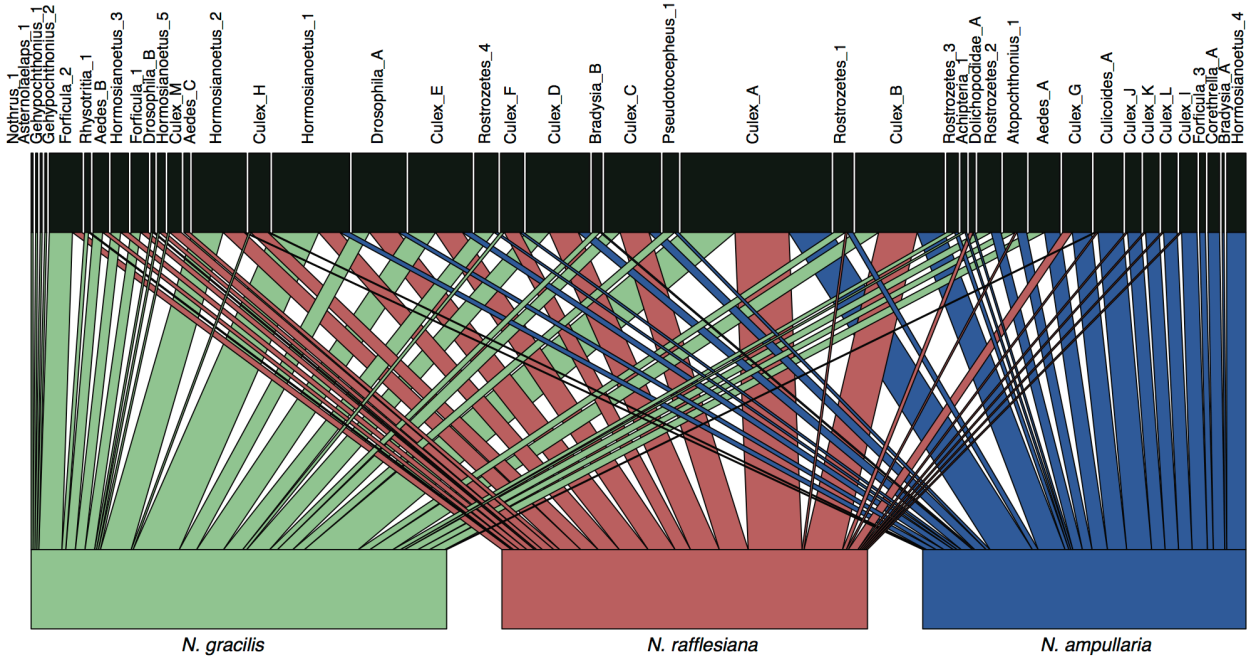


Figure 2.4:
Bipartite network of inquilines and *Nepenthes* host species. Inquiline OTUs are listed above, dipteran names are followed by letters and mites names are followed by numbers. The graph is organized to have the fewest crossing lines; OTUs associated with only one host species are towards the edges, while shared OTUs are in the center. Observations of 3 or fewer sequences were removed before calculating the presence or absence of each OTU in a sample. Line thickness is proportional to the number of times an inquiline OTU is found within a host species.

Results

18S quality control and OTU picking

MiSeq sequencing of samples generated a total of 5,501,913 eukaryotic sequences, which clustered into 23,444 OTUs. Of these, 14,302 OTUs were each represented by a single sequence, a common finding with NGS (Huse *et al.* 2010). "Singleton" OTUs may represent real observations of rare organisms, or be caused by sequencing errors. Numbers of sequences per sample ranged from 25,494 to 229,682. When arthropod OTUs were filtered and collected into a separate pool, the pool included 2,620,598 arthropod sequences, clustered into 7,229 OTUs (3,115 were singletons). Per sample sequence counts for arthropods ranged from 1,595 to 65,604.

Rarefaction curves for each sample of all observed eukaryotic OTUs plotted against sequences per sample appear to be leveling off (Figure 2.1C), suggesting sufficient sampling depth in our study, although for approximately 20% of the samples curves were still increasing at the cut-off of 25,494 sequences per sample. The average observed species per sample was around 300 eukaryotic OTUs (Figure 2.1C), with many OTUs observed only once. Ne72 appeared to have soil inside when collected. It possesses a much higher observed species richness (Figure 2.1C) and a different taxonomic composition than all other samples, and was excluded from subsequent analyses.

Arthropods are the most abundant organisms in the data, accounting for over 50% of the observed OTUs. The majority of arthropod sequences are insects, accounting for over 40% of observed OTUs (Figure 2.1A). The most common arthropods in samples are dipteran insects and mites, reflecting what are known to be the most abundant living arthropods in *Nepenthes*. Other common organisms found in samples are protists (algae, amoebae, ciliates, rotifers and others), fungi (mainly yeasts in the Saccharomycetes), and annelids (Figure 2.1A, Supplementary Table

2.2). Surprisingly, few OTUs were identified as ants, suggesting the DNA of most prey species was degraded before sampling.

Over 370,000 sequences were identified as gregarine protists (Gregarinasina). Gregarines are parasites of invertebrates, and in these samples the parasites emerge as the fourth most abundant group of eukaryotes in pitchers (Figure 2.1A, Supplementary Table 2.2). Gregarines may be living in the intestines of invertebrates in the pitchers, or may persist as free-living, infective sporozoites, presumably searching for new hosts in the pitcher microhabitats.

COI barcoding, and phylogenetic trees of COI and 18S

COI taxonomy and tree:

After quality control and removal of identical sequences, twenty-three unique COI barcodes of dipteran inquilines were assigned taxonomy (Supplementary Table 2.3) and visualized in a phylogenetic tree (Figure 2.2). GenBank accession numbers for the sequences are KP845038-KP845060. Mosquito larvae are the most abundant and diverse dipteran inquilines found in our tree. However, taxonomic assignment of the mosquito sequences was often poor, with many sequences having only 88-89% similarity to the top BLASTn hit in NCBI (Supplementary Table 2.3). In the COI phylogenetic tree, one mosquito clade has a bootstrap value of 100%, indicating uniform support for the group, and it contains sequences assigned to the genera *Tripteroides* and *Ochlerotatus* (Figure 2.2). Another mosquito clade of *Culex* species is also strongly supported; however, the final two mosquito sequences (Armigeres_1 and Ochlerotatus_4) are in poorly-supported clades and the mosquitoes as a whole are non-monophyletic.

Three other clades, representing *Corethrella*, *Dasyhelea* and *Endonepenthia* morphospecies, are each monophyletic with 100% bootstrap support. The COI sequences representing the *Corethrella* morphospecies were taxonomically assigned to *Sergentomyia* (Psychodidae), with only 86% sequence similarity for the top BLAST hit. According to both the morphology of the *Corethrella* larvae found in pitchers, and placement of sequences in the phylogenetic tree, the COI taxonomy assignment is incorrect. Like *Corethrella*, COI taxonomy assignments for sequences representing the *Dasyhelea* morphospecies were quite poor – top BLAST results had 84-86% similarity to three different genera in two different families (both probably incorrect assignments). However, for the *Endonepenthia* morphospecies, the COI taxonomy assignments were all to the same family and matched the morphospecies designation (Phoridae), with genus unspecified and sequence similarities of 89-90% (Supplementary Table 2.3).

18S taxonomy and tree:

Twenty-four OTUs identified as dipteran insects each had over 50 sequences in our rarefied arthropod OTU table, and were used to build a phylogenetic tree. The mosquito OTUs formed a monophyletic clade with 82% bootstrap support, separated into four main clades, three of which had sequences assigned to the genus *Culex*, and one with sequences assigned to *Aedes* (Figure 2.2).

The *Corethrella* morphospecies was represented by one OTU, with taxonomic assignment to the genus *Corethrella* at 100% sequence similarity (Supplementary Table 2.3). In the 18S tree, the *Corethrella* OTU falls sister to the mosquito clade (Figure 2.2). Similarly, the *Dasyhelea* morphospecies was represented by one OTU assigned to *Culicoides* (a genus in the

same family as *Dasyhelea*: Ceratopogonidae), and is sister to the other Culicomorpha. Sequence similarity was lower for this taxonomy assignment, at 93%. The *Lestodiplosis* morphospecies was represented by two OTUs, both assigned to *Bradysia* in a clade with high bootstrap support. However, the clade falls within a different clade of OTUs thought to represent the *Endonepenthia* morphospecies (Figure 2.2). The OTUs thought to represent the *Endonepenthia* morphospecies were taxonomically assigned to *Drosophila* and unnamed species from two other families: Asilidae and Dolichopodidae, all at 95-98% sequence similarity (Supplementary Table 2.3).

Comparing COI and 18S:

The COI and 18S trees correspond well, with the same placement for four of the five dipteran insect morphospecies (Figure 2.2). The *Lestodiplosis* morphospecies is the most problematic taxon for both COI and 18S phylogenies, and its placement is likely incorrectly resolved in both trees. Inclusion of this taxon caused other clades to be non-monophyletic (mosquitoes in the COI tree, and *Endonepenthia* in the 18S tree). In the COI phylogenetic tree, Cecidimyiidae_1 has a very long branch. According to both the morphological characters of the barcoded insect and the taxonomic assignment of the sequence, it is not closely related to mosquitoes, and therefore is incorrectly placed in the COI tree. The corresponding 18S sequences representing the *Lestodiplosis* morphospecies were taxonomically assigned to the *Bradysia* genus, and fall within a different clade (representing the *Endonepenthia* morphospecies), causing the clade to be paraphyletic just like the COI mosquito clade (Figure 2.2).

Taxonomic assignments of COI and 18S sequences corresponded with morphospecies taxonomy for some insects but not for others (Table 2.3). If we assume the Kai Lok 1997 taxonomy places inquilines in the correct families, then 18S has better taxonomic assignment for *Corethrella* and *Dasyhelea*, while COI has better taxonomic assignment for *Endonepenthia* and *Lestodiplosis* (Table 2.3).

Table 2.3. Taxonomy assignments and correspondence between COI and 18S

Morphospecies name		**COI taxonomy assignment**		**18S taxonomy assignment**		**Correspondence**
Genus	**Family**	**Genus**	**Family**	**Genus**	**Family**	
Corethrella	Corethrellidae	*Sergentomyia*	Psychodidae	*Corethrella*	Corethrellidae	Morphospecies genus = 18S genus, COI = different family
Dasyhelea	Ceratopogonidae	*Drosophila, Lutzomyia, Phlebotomus*	Drosophilidae, Psychodidae	*Culicoides*	Ceratopogonidae	Morphospecies family = 18S family, COI = different family
Endonepenthia	Phoridae	Phoridae sp.	Phoridae	*Drosophila,* Dolichopodidae sp., Asilidae sp.	Drosophilidae, Dolichopodidae, Asilidae	Morphospecies family = COI family, 18S = different families. All Muscomorpha infraorder.
Lestodiplosis	Cecidomyiidae	Cecidomyiidae sp.	Cecidomyiidae	*Bradysia*	Sciaridae	Morphospecies family = COI family, 18S different families. All Sciaroidea superfamily.
NA ("Mosquito")	Culicidae	*Culex, Ochlerotatus, Tripteroides*	Culicidae	*Aedes, Culex*	Culicidae	All same family.

Arthropod counts and comparisons with 18S sequences

Individual counts of inquilines correlate with 18S sequence counts, although the variance explained by regressions is low (Figure 2.3A). Most correlations are significant. The strongest correlation is for the *Dasyhelea* morphospecies: $R^2 = 0.80$, P = 1.03e-08. Other correlations, of mosquitoes, mites, and *Endonepenthia*, are weaker: R^2 values range from 0.22 to 0.38, and P-values range from 0.03 to 0.002. Counts of the *Lestodiplosis* morphospecies are not significantly correlated with sequence counts: $R^2 = 0.14$, P = 0.08. For non-normal inquiline distributions, the permutational linear models always agreed with the linear models in terms of significance, and we report only the results of the linear models for consistency. The *Corethrella* morphospecies was not included in this analysis because for the subset of 2012 samples counted, it was observed in only one sample (Table 2.3). Nevertheless, the morphospecies count matched the sequence data in terms of presence and absence, as the Corethrella_A OTU was only present in the same sample as the counted *Corethrella* insect (and the sequence was absent from the other samples in the subset).

The proportional rank-abundance plot shows similar curves for sequence and count data; however, ranks are not the same for inquilines of intermediate abundances (Figure 2.3B). Mites, *Dasyhelea*, and *Endonepenthia* switch ranks because the proportion of mite sequences is higher than the proportion of mite counts, while count proportions are higher than sequence proportions for *Dasyhelea* and *Endonepenthia* morphospecies.

Multivariate analyses

Diversity analyses using the 18S data showed highly significant differences among the arthropod communities by host species; with PERMANOVA: $R^2 = 0.098$, P = 0.001, and with

ANOSIM: R = 0.179, P = 0.001. Assumptions of the tests were not violated, as betadispersion of host species groups were not significantly different, P = 0.772. Finally, the pooling of *N. gracilis* samples did not have a significant effect, PERMANOVA: $R^2 = 0.029$, P = 0.135, and ANOSIM: R = 0.027, P = 0.255.

Network analysis

The bipartite network of arthropod inquilines and *Nepenthes* species (Figure 2.4) illustrates how certain OTUs are present in samples from only one host species (for example, Hormosianoetus_4, a mite, is only in *N. ampullaria* samples), while others are commonly found in samples from all three hosts (e.g. Culex_A). The network evaluates presence and absence (of all observations with more than three sequences), and the width of the line is proportional to the number of samples from a host species containing a particular OTU (Figure 2.4). The network has a relatively low level of specialisation: $H_2' = 0.26$, but it is highly significant when compared to a null model: $P < 0.001$. Our results are robust, as we found all networks to be significantly specialised, even when we did not correct for potential biases from very high or low abundance observations. A network with the same thresholds that included all OTUs from the rarefied arthropod OTU table had a very similar level of specialisation as the network of inquilines alone: $H_2' = 0.25$, $P < 0.001$.

A spring-embedded network (Supplementary Figure 2.1) reveals a core set of OTUs associated with all three *Nepenthes* host species. These 15 OTUs were identified as six mosquitoes, four mites, three wasps, two ants, and one fly. The fly OTU, named Drosophila_A, is in the *Endonepenthia* morphospecies clade (Figure 2.2). The ant OTUs were taxonomically

assigned to the genera *Leptothorax* and *Solenopsis*, and are likely to be the most common prey items in the sampled *Nepenthes* pitchers.

Discussion:

Is metabarcoding with 18S primers an effective tool for the characterisation of eukaryotic communities, and specifically arthropods?

Despite the limitations of 18S rRNA NGS metabarcoding, which include copy number variation, PCR bias, and short sequences, metabarcoding is an effective tool for characterising eukaryotic communities within pitcher plants; especially communities with arthropods. Insects and mites were the most highly represented organisms in our samples (Figure 2.1A).

Many of the organisms uncovered using NGS, including the algae, yeasts, amoebae, and ciliates, would be difficult to observe and count even with a microscope. Although the most abundant taxa would be readily recorded, the sampling effort required to capture the less abundant taxa would be prohibitive. Furthermore, parasitic life forms, such as the gregarine protists that accounted for over 6% of our sequences (Figure 2.1A, Supplementary Table 2.2), are likely embedded within other organisms and would not be detected by eye. The omission of organisms like gregarines from a dataset is a serious concern, because parasites play key roles within a community, by exerting selective pressures on hosts, controlling population sizes, and altering the structure of ecological networks (Hatcher *et al.* 2012). Similarly, fungi and protists are likely to be essential components of pitcher plant food webs, and in separate work, we are more exhaustively analysing data of these organisms (L. S. Bittleston, unpublished results). NGS provides a molecular window through which to observe buried or microscopic organisms.

Perhaps surprisingly, the majority of our sequences were from insect DNA (Figure 2.1A, Supplementary Table 2), but not ants. Instead, most sequences were from the inquiline insects living within *Nepenthes* pitchers. DNA may be largely degraded and decomposed in all but the living inquilines and the most recently captured prey. Our method of preservation likely facilitated the entry of inquiline DNA into sample fluid, as CTAB contains salts and detergent capable of lysing cells and releasing DNA. In fact, DNA of preserved insects has even been amplified and sequenced directly from the ethanol used to store samples (Shokralla *et al.* 2010), indicating DNA can readily be recovered from the fluid surrounding specimens.

How does 18S metabarcoding compare to COI Sanger sequencing in terms of taxonomic resolution?

We recommend that future studies of arthropod diversity in environmental samples use 18S metabarcoding combined with longer COI sequences wherever possible, as data from the two markers together will provide the best representation of communities. As sequence databases continue to expand, we expect taxonomic assignments from both 18S and COI will improve dramatically.

Our comparison of 18S amplicon sequencing with classic COI barcoding helped to "ground-truth" our sequencing approach in the sense that inquiline phylogenies made with these two markers largely agreed with each other. However, the comparison of the 18S and COI data underscored how taxonomic assignment of Southeast Asian insects is severely limited by databases. Previous studies have drawn attention to the same issue for insects from China (Yu *et al.* 2012). The much longer sequences and improved taxonomic resolution of COI barcodes should provide significantly better assignment than 18S OTUs; however, this was not the case in

our study. Although names given to *Nepenthes* inquiline morphospecies may not be exact, they are most likely in the correct family, as taxonomists have reared the larvae to adulthood and named them based on clear characters. For two dipteran inquilines, COI sequences BLASTed to the same family as the morphospecies name and 18S did not, and for two others the opposite was true (Table 2.3). Percent identity of both 18S and COI sequences compared to the best BLASTn hit is often very low, with the lowest being 83% for 18S and 84% for COI (Supplementary Table 2.3). We had expected COI barcodes to provide greater taxonomic resolution than the 18S amplicons, and similarly we expected the longer COI reads to build a better phylogenetic tree. However, this was also not the case. The COI and 18S phylogenetic trees are largely similar, and both have strong bootstrap support for most inquiline morphospecies clades (Figure 2.2). In both trees, *Lestodiplosis* is problematic, causing other clades to be non-monophyletic. The inquiline morphospecies with the most consistent assignment is the mosquito group: Culicidae is the only inquiline morphospecies family assigned by both 18S and COI (Table 2.3). The correspondence is likely due to the fact that Culicidae is well represented in databases, because genera such as *Aedes* and *Culex* are important vectors of human diseases, and thus have been sequenced extensively.

Are numbers of 18S sequences roughly representative of the number of individuals present in the communities?

The answer to this is a tentative yes in the sense that we found correlations between whole arthropod counts and scaled 18S sequences, although correlations are stronger for some organisms and weaker for others, and rank abundances do not correspond for inquilines of intermediate abundance (Figure 2.3). The difference between counts and sequences likely has

multiple causes. One potential cause is the amplicon sequencing process: copy number variation and both PCR and primer bias. Another difficulty is with counting all of the organisms in a sample, as they can be at different life history stages. Eggs and small larvae of dipterans or mites may pass through the sterile gauze we used to separate macrofauna from the pitcher fluid, but they would still be sequenced and would generate higher sequence abundances than counts. A third potential cause of variation is difficulty in assigning appropriate OTUs to individual inquilines. For example, the OTU representing the *Dasyhelea* morphospecies was easy to assign, as only one highly abundant OTU matched the morphospecies both in taxonomic assignment and placement in the phylogenetic tree. OTUs representing the *Lestodiplosis* morphospecies were more difficult to assign, as diversity in the sequences led to multiple OTUs and poor taxonomic assignments. A fourth source of variation is low sample size, as certain inquilines are rarer than others. For example, *Corethrella* was present in only one of our counted samples and could not be used to fit a model, and only four samples had non-zero individual counts of *Lestodiplosis* (Table 2.2). Finally, 18S sequence counts are only proxies for relative abundance, owing to standardization prior to sequencing, but arthropod counts may reflect variation in both relative and absolute abundance. Nevertheless, despite the different causes of variation between individual counts and sequence abundance, our results indicate that on the whole, 18S amplicon sequences roughly correspond to real counts and can be used to investigate community composition and structure, at least for the organisms we examined. At the moment, it is a useful method for uncovering the diversity of these microcosms with some reasonable indication of relative abundance.

Can 18S metabarcoding be used to study ecological dynamics and host specialisation in natural communities?

Our results show that 18S rRNA metabarcoding is currently an effective tool for studying ecological dynamics in *Nepenthes* pitcher plants. The method can be applied to other systems and questions, and is a viable option for more extensive studies.

Network analyses reveal specialisation of certain inquilines to particular hosts, as well as the presence of core inhabitants in all three *Nepenthes* species (Figure 2.4). The bipartite network of arthropods in *Nepenthes* pitcher microcosms is significantly specialised across the three host species, according to null models of the H_2' network-level specialisation index. Certain associations seem quite stable; for example, the OTU representing the *Dasyhelea* morphospecies was abundant in 75% of the sampled *N. ampullaria* pitchers, across three different locations and two years. In contrast, it was present in low levels in only 13% and 10% of *N. gracilis* and *N. rafflesiana* samples, respectively. Diversity analyses using multivariate statistics indicate arthropod communities are significantly different among host species, supporting the bipartite network results. These differences are also reflected in the spring-embedded network, where OTUs shared by samples from the same host species tend to cluster together (Supplementary Figure 2.1). In future studies it will be fascinating to see if core inhabitants of the *Nepenthes* species are found in other small aquatic habitats or only in pitcher plant microcosms.

Nepenthes species inhabiting the same relatively disturbed habitats in Singapore are nevertheless differentiated in their fauna, and as such seem to occupy (and construct) distinct ecological niches. Ongoing studies will expand the analysis of *Nepenthes* pitcher plant inhabitants to different species in more pristine habitats, and to other organisms including bacteria (L. S. Bittleston, unpublished results). This will further illuminate degrees of

specialisation and help uncover potential coevolution of organisms within *Nepenthes* microcosms.

Data and R code used in this study are available from the Harvard Dataverse Network. DOI:10.7910/DVN/27937

Chapter 2 Acknowledgements:

Our research was supported by an NSF Graduate Research Predoctoral Fellowship and an NSF Doctoral Dissertation Improvement Grant to L.S.B., and a Templeton FQEB Grant to N.E.P. and A.P. We thank Singapore National Parks for issuing collecting permits (Research Permit: NP/RP11-096-1 to L.S.B.), and the many people who provided suggestions and assistance with fieldwork, including Matthew Lim, Shawn Lum, Stuart Davies, Kang Min Ngo, Su Shiyu, and Kadeem Gilbert. Li Daiquin and Gregory Jedd graciously provided laboratory space in Singapore. We appreciate Charles Wolock's help with DNA extractions and inquiline taxonomic assignments, and Jon Sanders' support with sequencing preparation. We thank two anonymous reviewers for their suggestions that greatly improved the manuscript. This work was inspired in part by Roger Kitching's seminal contributions to the study of phytotelmata, and we are deeply grateful to him for his unflagging enthusiasm, generosity and intellectual leadership.

Chapter 2 References:

Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Applied and environmental microbiology 71:8966–8969.

Adlassnig, W., M. Peroutka, and T. Lendl. 2011. Traps of carnivorous pitcher plants as a habitat: composition of the fluid, biodiversity and mutualistic activities. Annals of botany 107:181–194.

Albert, V. A., S. E. Williams, and M. W. Chase. 1992. Carnivorous plants: phylogeny and structural evolution. Science 257:1491–1495.

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

Amaral-Zettler, L. A., E. A. McCliment, H. W. Ducklow, and S. M. Huse. 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PloS one 4:e6372.

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. Austral ecology 26:32–46.

Anderson, M. J. 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics 62:245–253.

Basset, Y., L. Cizek, P. Cuénoud, R. K. Didham, F. Guilhaumon, O. Missa, V. Novotny, F. Ødegaard, T. Roslin, J. Schmidl, A. K. Tishechkin, N. N. Winchester, D. W. Roubik, H.-P. Aberlenc, J. Bail, H. Barrios, J. R. Bridle, G. Castaño-Meneses, B. Corbara, G. Curletti, W. Duarte da Rocha, D. De Bakker, J. H. C. Delabie, A. Dejean, L. L. Fagan, A. Floren, R. L. Kitching, E. Medianero, S. E. Miller, E. Gama de Oliveira, J. Orivel, M. Pollet, M. Rapp, S. P. Ribeiro, Y. Roisin, J. B. Schmidt, L. Sørensen, and M. Leponce. 2012. Arthropod diversity in a tropical forest. Science 338:1481–1484.

Beaver, R. A. 1985. Geographical variation in food web structure in *Nepenthes* pitcher plants. Ecological Entomology 10:241–248.

Bik, H. M., D. L. Porazinska, S. Creer, J. G. Caporaso, R. Knight, and W. K. Thomas. 2012. Sequencing our way towards understanding global eukaryotic biodiversity. Trends in ecology & evolution 27:233–243.

Blüthgen, N., F. Menzel, and N. Blüthgen. 2006. Measuring specialization in species interaction networks. BMC ecology 6:9.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences of the United States of America 108 Suppl 1:4516–4522.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, and M. Bauer. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal 6:1621–1624.

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

Chapin, C. T., and J. Pastor. 1995. Nutrient limitations in the northern pitcher plant *Sarracenia purpurea*. Canadian Journal of Botany 73:728–734.

Clarke, C. M., and R. L. Kitching. 1993. The metazoan food webs from six Bornean *Nepenthes* species. Ecological Entomology 18:7–16.

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. Austral ecology 18:117–143.

Deagle, B. E., S. N. Jarman, E. Coissac, F. Pompanon, and P. Taberlet. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biology letters 10.

Dormann, C. F. 2011. How to be a specialist? Quantifying specialisation in pollination networks. Network Biology 1:1–20.

Dormann, C. F., J. Fründ, N. Blüthgen, and B. Gruber. 2009. Indices, graphs and null models: analyzing bipartite ecological networks. The Open Ecology Journal 2:7–24.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32:1792–1797.

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

Ellison, A. M., and N. J. Gotelli. 2009. Energetics and the evolution of carnivorous plants—Darwin's "most wonderful plants in the world." Journal of Experimental Botany 60:19–42.

Fashing, N. J. 2002. *Nepenthacarus*, a new genus of Histiostomatidae (Acari: Astigmata) inhabiting the pitchers of *Nepenthes mirabili*s (Lour.) Druce in far north Queensland, Australia. Australian Journal of Entomology 41:7–17.

Fashing, N. J., and T. H. Chua. 2002. Systematics and ecology of *Naiadacarus nepenthicola*, a new species of Acaridae (Acari: Astigmata) inhabiting the pitchers of *Nepenthes bicalcarata* Hook. F. in Brunei Darussalam. International Journal of Acarology 28:157–167.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Molecular marine biology and biotechnology 3:294–299.

Gronemeyer, T., F. Coritico, A. Wistuba, D. Marwinski, T. Gieray, M. Micheler, F. S. Mey, and V. Amoroso. 2014. Four new species of *Nepenthes* L. (Nepenthaceae) from the central mountains of Mindanao, Philippines. Plants 3:284–303.

Hajibabaei, M., S. Shokralla, X. Zhou, G. A. C. Singer, and D. J. Baird. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PloS one 6:e17497.

Hatcher, M. J., J. T. A. Dick, and A. M. Dunn. 2012. Diverse effects of parasites in ecosystems: linking interdependent processes. Frontiers in Ecology and the Environment 10:186–194.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270:313–321.

Huse, S. M., D. M. Welch, H. G. Morrison, and M. L. Sogin. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental microbiology … 12:1889–1898.

Kai Lok, C., L. M. J. Chen, J. P. S. Choo, K. T. Ling, L. T. Ming, P. K. L. Ng, and T. W. Kiatt. 1997. A guide to the carnivorous plants of Singapore. (H. T. W. Tan, Ed.). Singapore Science Centre, Singapore.

Kitching, R. L. 2000. Food webs and container habitats. Cambridge University Press.

Klappenbach, J. A., P. R. Saxman, J. R. Cole, and T. M. Schmidt. 2001. rrndb: the Ribosomal RNA Operon Copy Number Database. Nucleic acids research 29:181–184.

Kneitel, J. M., and T. E. Miller. 2002. Resource and top-predator regulation in the pitcher plant (*Sarracenia purpurea*) inquiline community. Ecology 83:680–688.

Mogi, M., and H. S. Yong. 1992. Aquatic arthropod communities in *Nepenthes* pitchers: the role of niche differentiation, aggregation, predation and competition in community organization. Oecologia 90:172–184.

Moran, J. A., W. E. Booth, and J. K. Charles. 1999. Aspects of pitcher morphology and spectral characteristics of six Bornean *Nepenthes* pitcher plant species: implications for prey capture. Annals of botany 83:521–528.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2013, December. vegan: Community Ecology Package.

Prokopowich, C. D., T. R. Gregory, and T. J. Crease. 2003. The correlation between rDNA copy

number and genome size in eukaryotes. Genome / National Research Council Canada = Génome / Conseil national de recherches Canada 46:48–50.

Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glöckner. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic acids research 35:7188–7196.

R Core Team. 2014, January 28. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ratsirarson, J., and J. A. Silander. 1996. Structure and dynamics in *Nepenthes madagascariensis* pitcher plant micro-communities. Biotropica 28:218–227.

Riley, C. V. 1874. On the insects more particularly associated with *Sarracenia variolaris* (Spotted Trumpet-Leaf.). The Canadian Entomologist 6:207–214.

Shokralla, S., G. A. C. Singer, and M. Hajibabaei. 2010. Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. BioTechniques 48:233–234.

Sodhi, N. S., L. P. Koh, B. W. Brook, and P. K. L. Ng. 2004. Southeast Asian biodiversity: an impending disaster. Trends in ecology & evolution 19:654–660.

Srivastava, D. S., J. Kolasa, J. Bengtsson, A. Gonzalez, S. P. Lawler, T. E. Miller, P. Munguia, T. Romanuk, D. C. Schneider, and M. K. Trzcinski. 2004. Are natural microcosms useful model systems for ecology? Trends in ecology & evolution 19:379–384.

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

Stoeck, T., D. Bass, M. Nebel, R. Christen, M. D. M. Jones, H.-W. Breiner, and T. A. Richards. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology 19 Suppl 1:21–31.

Taberlet, P., E. Coissac, F. Pompanon, C. Brochmann, and E. Willerslev. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21:2045–2050.

Yu, D. W., Y. Ji, B. C. Emerson, X. Wang, C. Ye, C. Yang, and Z. Ding. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution 3:613–623.

Chapter 3: Convergence among the microbiomes of Southeast Asian and North American pitcher plants

Leonora S. Bittleston[a], Charles Wolock[a], Effendi Y. Bakhtiar[b], Xin Yue Chan[c], Kok Gan Chan[c], Naomi E. Pierce[a], Anne Pringle[d]

[a]Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, [b]Institute for Tropical Biology and Conservation, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia, [c]High Impact Research Institute, University of Malaya, Kuala Lumpur, Malaysia, [d]Departments of Botany and Bacteriology, University of Wisconsin-Madison, Madison

Corresponding author: Leonora S. Bittleston; address: 26 Oxford St., Cambridge MA 02143; phone:1-831-212-1936; email: leobit@gmail.com

Abstract:

Recognition of convergently evolved interactions is likely to enable identification of similar selective pressures shaping microbial communities, even across distant parts of the world. Pitchers of carnivorous plants are exquisite examples of convergent evolution; pitcher morphologies have evolved independently within three plant orders. Pitchers trap and digest prey but also house complex communities of associated organisms. An open question is whether communities within pitchers will also show signals of convergence. We sequenced fluids from

more than 330 wild pitchers of eight Southeast Asian *Nepenthes* species and six American *Sarracenia* species, and find evidence for convergence of pitcher microbiomes. Compared to surrounding habitats, pitcher communities have lower species richness and evenness, and while the communities of the two pitcher systems are made up of different species, the communities appear to be phylogenetically related subsets of the bacteria and eukaryotes found in nearby soil or bog water. Moreover, when Southeast Asian *Nepenthes* are placed in a North American bog, *Nepenthes* pitchers assemble *Sarracenia*-like communities. Tubes with a pitcher-like form also assemble similar communities, suggesting pitcher shape and structure influences microbial colonization. Within each genus of pitcher plants, pitcher characteristics drive between-species diversity; *Nepenthes* microbiome diversity is shaped by differences in pH while *Sarracenia* microbiome diversity is influenced by pitcher morphology. In an epoch characterized by both biodiversity discovery and loss, it is essential to understand more about how species interactions structure ecosystems. Convergent interactions can be used to accelerate biodiversity discovery, particularly of microbial communities.

Significance Statement:

The aquatic pools held within pitchers of carnivorous pitcher plants represent entire small ecosystems. Pitcher plant form and function has evolved independently multiple times, and thus these systems allow us to examine how evolution has shaped the communities living within pitchers in similar or different ways. We demonstrate convergence in species richness, evenness, phylogenetic composition, and a functional trait among the microbiomes of multiple species from two distantly related pitcher plant genera from Southeast Asia and North America. This research helps to uncover consistent patterns in the formation of complex communities on

opposite sides of the globe.

Introduction:

Similar selective pressures in geographically distant habitats can cause unrelated organisms to converge in both morphological and functional traits. Pitchers of carnivorous plants are exquisite examples of convergent evolution: pitcher plants have evolved repeatedly and independently in Southeast Asia, North America and Australia (Albert et al. 1992). Similar selective pressures can also cause the independent emergence of multispecies interactions with similar physiological or ecological functions across the different systems, defined as convergent interactions (Bittleston et al. 2016). Therefore, the ecological communities formed within unrelated pitchers might exhibit convergent structure. The communities living within the plant-held waters (phytotelmata) of carnivorous pitcher plants have long provided models of interspecific interactions and community dynamics because they are discrete, naturally replicated, and experimentally tractable. We use pitchers of convergently evolved genera of pitcher plants to test the nascent concept of convergent interactions.

Microbes comprise the majority of Earth's biodiversity. While centuries of natural history form the basis of animal and plant biogeography, the relatively recent advent of molecular tools, including next generation sequencing technologies, are only beginning to elucidate global patterns of microbial biodiversity (Nemergut et al. 2011, Bates et al. 2012, Gibbons et al. 2013, Davison et al. 2015). Global change is causing rapid losses of animal biodiversity (Dirzo et al. 2014), but while sufficient data are available to predict for example how global changes in climate and land-use will affect the 8,750 species of land birds (Jetz et al. 2007), we know relatively little about if or how global change will impact microbial communities, as there are

few baseline data from which to make estimates of loss or change.

To accelerate discovery of microbial biodiversity, new approaches are needed, and understanding the extent to which convergent morphologies trigger the formation of convergent microbiomes may facilitate the rapid collection of baseline data (Bittleston et al. 2016). Pitchers are elegant systems for exploring host-community dynamics, but pitcher plant fluids can also be thought of as discrete, aquatic ecosystems ideal for testing principles of microbiome assembly.

The genera *Sarracenia* and *Nepenthes* evolved independently in North America and Southeast Asia, respectively (Albert et al. 1992). Both genera grow pitchers to attract, trap, and digest insects (primarily ants), but pitchers also house communities of bacteria, fungi, protozoa, and arthropods (Beaver 1983, Bradshaw and Creelman 1984, Kitching 2000). Pitchers are similarly shaped, and pitchers of both genera both offer extra-floral nectar to attract prey, possess slippery interiors to trap prey, and secrete digestive enzymes to break down prey tissues (Juniper et al. 1989, Adlassnig et al. 2011, Kurup et al. 2013). Pitchers actively absorb nitrogen, phosphorus, and other elements from prey; these nutrients are otherwise scarce in the soils where the plants grow (Chapin and Pastor 1995, Schulze et al. 1997, Ellison 2006). Pitcher interiors appear to be sterile before opening (Peterson et al. 2008, Buch et al. 2013) (but see (Chou et al. 2014)), and once open, a complex community forms *de novo* within the pitcher (Beaver 1983, Bradshaw and Creelman 1984, Clarke and Kitching 1993, Kitching 2000, Kneitel and Miller 2002, Peterson et al. 2008, Koopman et al. 2010, Krieger and Kourtev 2012, Chou et al. 2014, Bittleston et al. 2015, Chan et al. 2016, Sickel et al. 2016). While the plants are perennial, individual pitchers last from a few weeks to two years, and are generally most active for the first few months after opening (Heard 1998, Osunkoya et al. 2008).

After collecting from across the Eastern Seaboard of the United States, and in Singapore and Malaysian Borneo, we sequenced phytotelmata from over 330 wild pitchers of *Sarracenia* and *Nepenthes* species in their natural habitats to characterize the biodiversity housed in pitcher microcosms. We tested for convergence of *Sarracenia* and *Nepenthes* microbiomes by comparing the species richness, community composition, phylogenetic structure, and functional potential of the different communities. We characterized both bacterial and eukaryotic communities for 14 different species of pitcher plants, and explored which features of host species best predict biodiversity. Finally, in a manipulative experiment, we tested whether North American insects and microbes would colonize Southeast Asian *Nepenthes* if *Nepenthes* plants were placed in a North American *Sarracenia* habitat. We discovered evidence for convergence of unrelated pitcher communities, and suggest aspects of pitcher form and physiology that may drive this similarity.

Materials and Methods:

Collections and experiments took place from 2012 to 2014 at field sites along the American Gulf Coast, at Harvard Forest in Massachusetts, in Singapore, and in the Maliau Basin of Borneo (Figure 3.1A). In total, we sampled and sequenced the communities of more than 330 wild pitchers, 70 environmental samples, 60 experimental pitchers, and 16 experimental tubes (for samples and metadata see Supplementary Table 3.1).

Collecting in undisturbed habitats:

Nepenthes pitchers from three co-occuring species (*N. ampullaria*, *N. gracilis* and *N. rafflesiana*) were sampled from three sites in Singapore (Kent Ridge Park, Bukit Timah Nature

Preserve, and between Lower and Upper Peirce Reservoir Park) in January 2012. Additional pitchers from the same species and sites were sampled in March 2013 and March 2014. Pitchers from an additional five co-occurring species (*N. veitchii, N. tentaculata, N. stenophylla, N. reintwardiana,* and *N. hirsuta*) were sampled from the Maliau Basin, Borneo in March 2014.

Sarracenia pitchers from five species (*S. alata, S. flava, S. leucophylla, S. rosea* and *S. rubra*) were sampled from thirteen sites along the American Gulf Coast in June 2014 and a sixth species (*S. purpurea*) was sampled from Harvard Forest in Massachusetts in July 2014. Gulf Coast species and sites included: in Mississippi (*S. alata* only)- Sweet Bay Bogs Preserve, Buttercup Flats, and Old Fort Bayou; in Alabama (*S. alata, S. leucophylla*, *S. rosea* and *S. rubra*)- Splinter Hill Bog and Week's Bay Pitcher Plant Bog; and in Florida (*S. flava*, *S. leucophylla,* and *S. rosea*)- Blackwater River State Forest, two sites on Nokuse Plantation, two sites on Eglin Airforce Base, and three sites in the Apalachicola National Forest. Sites were considered different if separated by more than 0.1 degree of latitude or longitude.

Sampling pitcher fluid:

Contents of each pitcher were collected with sterile, single-use plastic transfer pipettes and placed into empty, sterile plastic tubes. Fluids were mixed within each pitcher using the pipette before collecting to homogenize any differences by depth. Volumes and pH levels of all pitcher fluids were recorded, except for our first collection in Singapore in 2012 (for more detail on Singapore sampling see Bittleston et al. 2015). In some pitchers the volumes can be very large (e.g. 100 - 500 mLs); for higher volume samples we estimated total volume and collected a well-mixed subsample from the pitcher. We measured pH with colorpHast strips (EMD Millipore) by removing small amounts of fluid with additional sterile pipettes. To preserve DNA, we added

cetyl tri methyl ammonium bromide and salt solution (hereafter 'CTAB'; final concentrations: 2% CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris pH 8) to each sample in the same volume as the collected fluid. All samples were processed the same day as collection, except for Maliau Basin samples that were refrigerated overnight and processed the next morning, due to time constraints. After CTAB addition, samples were transported at room temperature to Harvard University, and subsequently frozen.

Sampling the surrounding environment:

Protocols reflected pitcher plant habitats (Figure 3.1A). When wet, we collected bog samples, and if dry, we collected soil samples and water either from fallen leaves or from sterile tubes placed in the environment, as follows: Singapore, March 2013—soil, Gulf Coast, June 2014—soil and bog water; Massachusetts, July 2014—bog water; Singapore, February 2014—sampling from plastic tubes left out for one month to collect rainwater and acquire microbial communities; Maliau Basin, January 2014—sampling from soil and water held in fallen leaves. All soil samples were collected from the surface organic layer. See Supplementary Table 3.1 for sample details.

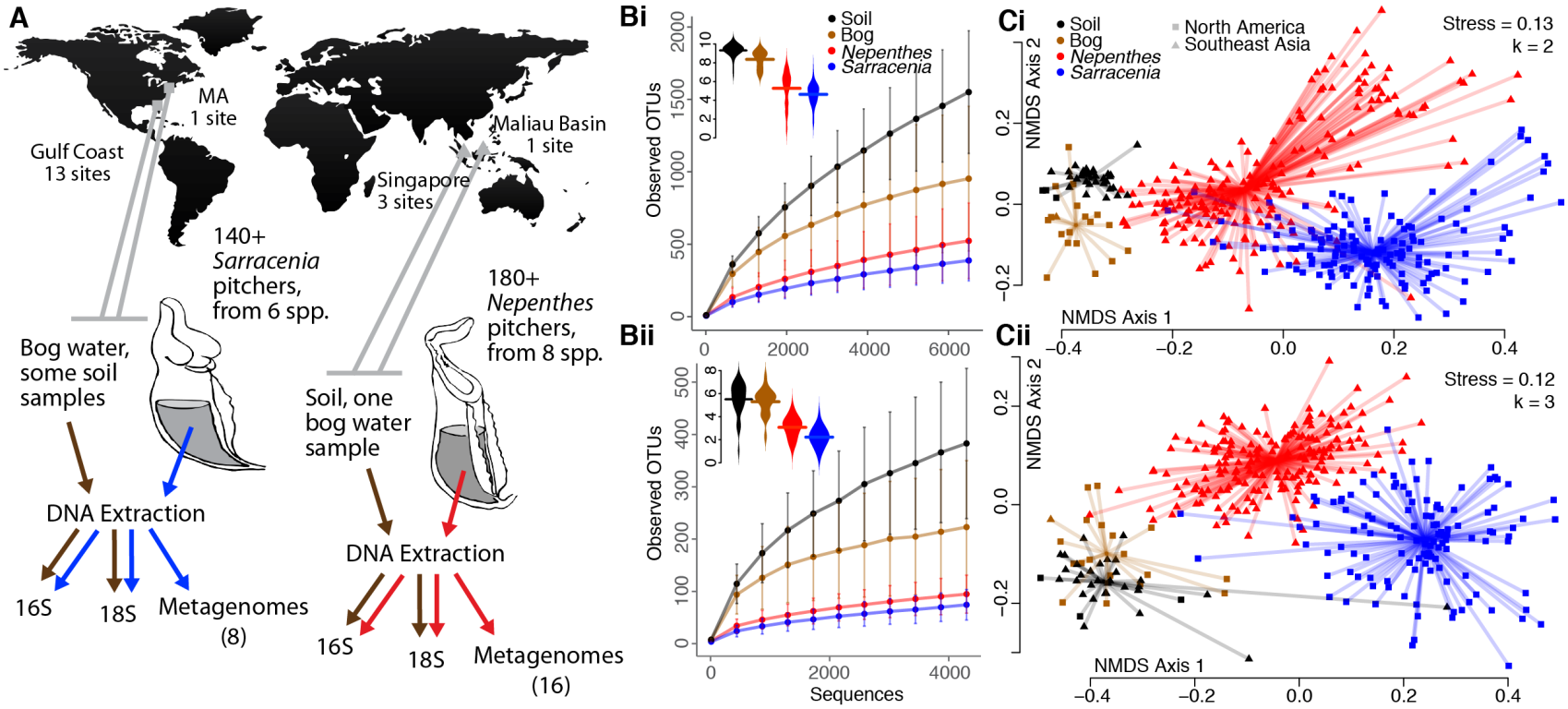


Figure 3.1: A) Sampled *Sarracenia* and *Nepenthes* natural populations. B) Species richness (displayed as rarefaction plots) and Shannon diversity (inset beanplots) of both bacterial (i, top) and eukaryotic (ii, bottom) communities was lower in pitchers than in surrounding soil and bog water. Error bars represent standard deviations. C) Community composition of bacteria (i, top) and eukaryotes (ii, bottom) differed between the two genera of pitcher plants and also differed from surrounding habitats. NMDS stress and dimensions (k) are listed, the center of each cluster is the category's median value.

Experimental relocation of *Nepenthes spp*. to a New England bog:

In summer 2013 we set up experiments to manipulate *Nepenthes* within a *Sarracenia* habitat, the Tom Swamp bog at Harvard Forest. Four different species of *Nepenthes* were sourced from Borneo Exotics via the ExoticPlantsPlus nursery in New York. *Nepenthes* plants were maintained in a greenhouse for two months after arriving from Southeast Asia, and then a growth chamber while pitchers were maturing for use in the field. *Sarracenia purpurea* plants were purchased from Meadowview Biological Research Station, potted using purchased sphagnum peat and perlite, maintained in a greenhouse for three months, and used in the experiments as a control for whether growth in a pot influenced community assembly. Three different experiments used different species and treatments based on the availability of experimental material, and were primarily driven by the number of plants with unopened pitchers per species.

After a preliminary experiment used to optimize protocols, Experiment II used 6 treatments: *S. purpurea* growing naturally in the bog, *S. purpurea* in pots, *N. ampullaria*, *N. gracilis* and *N. rafflesiana* in pots, and empty, 50 mL sterile glass tubes used as a rough mimic of the pitcher shape. Experiment III used 5 treatments: *S. purpurea* in the bog, *S. purpurea* in pots, *N. ampullaria* in pots, sterile glass tubes, and sterile glass tubes each filled with 30 mg of autoclaved, ground wasps as a nutrient and prey control. Experiment IV used 6 treatments: *S. purpurea* in the bog, *N. ampullaria*, *N. rafflesiana* and *N. bicalcarata* in pots, sterile glass tubes, and sterile glass tubes with prey.

Plants of Experiment II were grouped into 8 stations (9 – 16) and the experiment ran from June 26 – July 17. Pitchers were sampled for subsequent sequencing on days 14 and 21.

Experiment III (stations 17 – 21) ran from July 17 – Sept. 4. Pitchers were sampled for subsequent sequencing on days 14, 35, and 49. Experiment IV (stations 22 – 26) ran from July 24 – Sept. 10. Pitchers were sampled for subsequent sequencing on days 15, 35, and 48.

To sample, we collected 750 uL of fluid from experimental pitchers and tubes using sterile transfer pipettes, as described above. At sampling, we also noted the presence or absence of pitcher plant mosquito larvae (*Wyeomyia smithii*) in pitchers. On the last day of each experiment, we collected entire pitcher contents. On the last days of Experiments III and IV, we collected samples of bog water. All samples were stored in small tubes in a cooler with ice, and brought back to the laboratory, where they were frozen the same day. Subsequent analyses target only the last day's sample from each pitcher or tube, so that no pitcher or tube is included more than once in the dataset.

Sample processing, DNA extraction, amplification and sequencing:

Once our samples were collected, we turned our attention to DNA extraction and sequencing. During preparation of our first collections from Singapore in 2012, we filtered out macroscopic arthropods and prey with sterilized gauze before DNA extraction from fluids. For subsequent samples, we realized this was unnecessary, and just avoided macroscopic organisms when removing fluid for DNA extraction. To concentrate biological materials, half of the fluid from each sample was either filtered through sterilized 0.22-μm Durapore filters in Swinnex holders (Singapore 2012 and Experiment II), centrifuged (Singapore 2013), or 0.5 mL was removed from each sample, precipitated with an equal volume of isopropanol, and centrifuged (Maliau Basin, Gulf Coast, Singapore 2014 and Experiments III and IV). Protocols were optimized over the course of DNA extractions, and ultimately we targeted isopropanol

precipitation and centrifugation as the concentrating technique of choice because it required significantly less time and effort per sample. Subsequent analyses suggest sample concentration protocol has no impact on downstream sequencing (samples processed by any of the three techniques do not cluster together).

We extracted DNA by bead-beating concentrated materials with buffer and phenol-chloroform, and then proceeding with a standard phenol-chloroform extraction (Sambrook and Russell 2001). We included a negative control with each set of extractions. We measured DNA quantity, and then re-extracted DNA from a few samples with very low DNA amounts, using a larger initial volume. DNA extracts with dark coloration (suggesting high levels of polyphenols) were cleaned using a MoBio Powerclean kit. DNA of successful extractions was fluorometrically quantified a final time using a Quant-iT High-Sensitivity dsDNA Assay Kit (Invitrogen). A subset of samples was sent to Argonne National Laboratories for Illumina MiSeq next-generation amplicon sequencing in 2013 .The Earth Microbiome Project's barcoded 16S and 18S primers (Amaral-Zettler et al. 2009, Caporaso et al. 2012) were used to amplify DNA in separate runs, with PCR amplification and sequencing executed according to the Earth Microbiome Project protocols (http://www.earthmicrobiome.org/emp-standard-protocols). The 16S primers target the V4 region of the ribosomal RNA gene and are generally used to characterize bacterial communities, while the 18S primers target the V9 region and are used to characterize eukaryotes. We sequenced the rest of our samples in 2015, using the same approach.

Shotgun metagenomics:

To explore functional gene diversity in pitcher plant microbiomes, we conducted shotgun metagenomic sequencing with 24 of our pitcher samples. Sixteen samples sequenced at the High

Impact Research Institute in Malaysia (HIR), and eight at Harvard University. The HIR set targeted two samples from each of four different species of both *Nepenthes* and *Sarracenia*. DNA was extracted as described above, but each sample was extracted 2-4 times and extractions pooled to increase the amount of DNA available for sequencing. DNA was sheared at settings aiming for average lengths of 350 base pairs (bp), and libraries were prepared using a TruSeq DNA PCR Free HT Kit. The quality and quantity of the DNA libraries were tested with a Bioanalyzer, pooled in equal concentrations and sequenced on the Illumina HiSeq 2500 platform in four Rapid Run, paired-end,100 bp lanes.

The Harvard University set included eight additional *Nepenthes* samples (two samples from each of four different species). Here, we used the same DNA extractions as previously used for metabarcoding. DNA samples were sheared with a Covaris at 500 bp, prepared with a Kapa LTP Library Prep Kit. Due to lower initial DNA quantities, the samples were subject to 2-9 cycles of PCR before final quantification using a PerfeCta GGS Library Quantification Kit, quality testing with a Bioanalyzer, and pooling of samples in equal concentrations. These libraries were sequenced in one-third of an Illumina HiSeq 150 bp paired-end Rapid Run lane.

Analyses of 16S and 18S diversity:

To generate Operational Taxonomic Units (OTUs), amplicon data were clustered using QIIME (Quantitative Insights Into Microbial Ecology) versions 1.8 and 1.9 (Caporaso et al. 2010) on Harvard University's Odyssey computing cluster. We joined forward and reverse reads using fastq-join, then split libraries with a PHRED quality cut-off of 20 to remove low-quality sequences, and used UCLUST open-reference clustering to form groups of sequences into OTUs with 97% similarity. Phylogenetic trees were generated using QIIME default settings for 16S;

when generating the18S alignment and tree we set the allowed gap fraction to 0.8 and the entropy threshold to 0.0005. We assigned taxonomy with the greengenes (Greengenes Database Consortium) and SILVA databases for 16S and 18S, respectively. Very few sequences were assigned to Archaea in our 16S dataset, so we chose to refer to it as 'bacteria' in subsequent analyses. For 18S, we used the BLAST method to assign taxonomy, as UCLUST assignment was poor. For subsequent analyses of 18S sequences, we used only OTUs assigned to Eukaryota.

To calculate alpha diversity of our samples, we first discarded any samples with fewer than 6,500 or 4,000 sequences for the 16S and 18S datasets, respectively. We then built rarefaction plots of soil, bog, *Nepenthes*, and *Sarracenia* samples using the observed species metric in QIIME and standard deviations across each category (Figure 3.1B). We also calculated the Shannon diversity index and standard deviations across the same categories (Figure 3.1B, Supplementary Table 3.2).

To explore beta diversity among samples, we first removed any observation of an OTU with less than 10 sequences per sample in order to minimize the probability of including sequencing errors, including barcode misassignments (Bokulich et al. 2013, Nelson et al. 2014). We accounted for uneven sequencing across the samples by subsampling our OTU tables to 4,000 sequences per sample. We calculated dissimilarity matrices with the unweighted Unifrac metric (Lozupone and Knight 2005) using R packages *picante* and *phyloseq* (Kembel et al. 2010, McMurdie and Holmes 2013), and ran non-metric multidimensional scaling (NMDS) analyses using the *vegan* R package (Oksanen et al. 2013) (Figures 3.1C, 3.3,3. 4, 3.5 and Supplementary Figure 3.1). We used the functions *envfit* and *ordisurf* to fit environmental factors or vectors (respectively) to our ordinations and to analyze main effects. In addition, we used mantel tests to test for correlations of the dissimilarity matrices with pH and volume, and permutational

multivariate analyses of variance (PERMANOVAs, function *adonis* in *vegan*) to test the explanatory power of factors including plant species and collection site (Supplementary Table 3.3).

To examine phylogenetic patterns among *Nepenthes* pitchers, *Sarracenia* pitchers, and environmental samples, we chose to focus on relatively common OTUs, removing OTUs containing fewer than 100 sequences across all our samples as well as those not present in at least 10% of either *Nepenthes* or *Sarracenia* microbiome samples. We then subsampled the OTU table for each category to 2,000 sequences per sample, combined all observations of the OTUs by category (e.g. *Nepenthes*, *Sarracenia* or environment), and normalized by the number of samples in each category. We filtered our previously generated 16S and 18S phylogenetic trees using the resulting OTU tables and plotted them with the Interactive Tree of Life (iToL) program(Letunic and Bork 2011) (Figure 3.2). We added barcharts along the outer edge of the trees, displaying the natural log of the abundance for each OTU in each category (Figure 3.2).

Functional analyses:

For the shotgun metagenomic data, we combined forward and reverse reads from all lanes for each sample, and used Trimmomatic to remove barcodes and low-quality sequences. We used HUMAnN2 (HMP Unified Metabolic Analysis Network 2, (Abubucker et al. 2012)) on Harvard University's Odyssey computing cluster to identify individual reads by comparing and annotating reads to reads of known function, build profiles of identified functional genes for each sample, and normalize numbers of sequences across samples . We next uploaded our metagenomes to MG-RAST (Glass et al. 2010), where they will be publically accessible, and compared our metagenomes to publically available metagenomes from soil, lake, and

phyllosphere habitats using MG-RAST and the NCBI's Short Reads Archive (SRA) (Supplementary Table 3.4). We analyzed these metagenomes in the same way as our own data, also using HUMAnN2. We made NMDS plots of KO functional matrices with Bray-Curtis distances and a beanplot of chitinases using the *vegan* and *beanplot* (Kampstra 2008) packages in R. We tested for a difference in chitinase abundances between pitcher plant and comparison metagenomes using a Mann-Whitney U test (function *wilcox.test* in R).

Results

North American and Southeast Asian pitcher plant communities converge in richness and evenness

To compare the microbial communities of *Sarracenia* and *Nepenthes* pitchers, we analyzed DNA samples from pitchers and their surrounding environments with an amplicon sequencing approach for characterizing both bacteria and eukaryotes (Figure 3.1A, Table 3.1). *Nepenthes* and *Sarracenia* pitcher communities were distinct from and had fewer OTUs than those in surrounding environments (Figure 3.1B, 3.1C). This was true for both bacteria and eukaryotes. Pitchers housed fewer OTUs and had significantly lower Shannon diversity (Mann-Whitney U Test, $P < 0.001$ for both bacteria and eukaryotes) than soil or bog water, indicating decreased richness and evenness (Figure 3.1B, Supplementary Table 3.2). Community composition in pitchers was significantly different from that of surrounding soil or bog water (Figure 3.1C, Supplementary Table 3.3. Bacteria: *envfit*: $R^2= 0.31$, $P < 0.001$, *adonis*: $R^2= 0.08$, $P < 0.001$; Eukaryota: *envfit*: $R^2= 0.38$, $P < 0.001$, *adonis*: $R^2= 0.08$, $P < 0.001$). To understand community differences across just one region, we separately tested the community compositions of Southeast Asian samples from pitchers, soil, bog water, plastic tubes, or cupped, dead leaves

filled with water and sitting on the ground. We found that the community compositions in water from leaves or from plastic tubes were both similar to pitcher fluid, while those of soil or bog water were very distinct (Supplementary Figure 3.1).

Table 3.1: Numbers of amplicon sequences and OTUs

	Primer set	**16S**	**18S**
	Samples	492	472
All data (post quality control)	Sequences	12,375,647	14,664,713
	OTUs	139,549	34,236
Without observations of fewer than 10 sequences	Sequences	11,337,226	14,447,905
	OTUs	12,940	7,516

Organisms common to *Nepenthes* and *Sarracenia* tend to be from the same phylogenetic clades

To compare phylogenetic patterns in pitcher communities, we mapped Operational Taxonomic Units (OTUs, clustered at 97% sequence similarity as a proxy for species) present in at least 10% of our *Nepenthes* or *Sarracenia* samples onto bacterial and eukaryotic phylogenetic trees, along with OTUs found in bog water and soil (Figure 3.2). Organisms repeatedly colonizing *Nepenthes* or *Sarracenia* pitchers tended to be from similar clades of bacteria or eukaryotes (Figure 3.2). The pattern was most pronounced in bacteria, and the most obviously shared families included Microbacteriaceae, Gordoniaceae, Chitinophagaceae, Sphingobacteriaceae, Bradyrhizobiaceae, Rhizobiaceae, Sphingomonadaceae, Burkholderiaceae, Comamonadaceae, Oxalobacteriaceae, Neisseriaceae, Enterobacteriaceae, Moraxellaceae, and Xanthomonadaceae. For eukaryotes the most obviously shared clades included dipteran insects, mites, and rotifers.

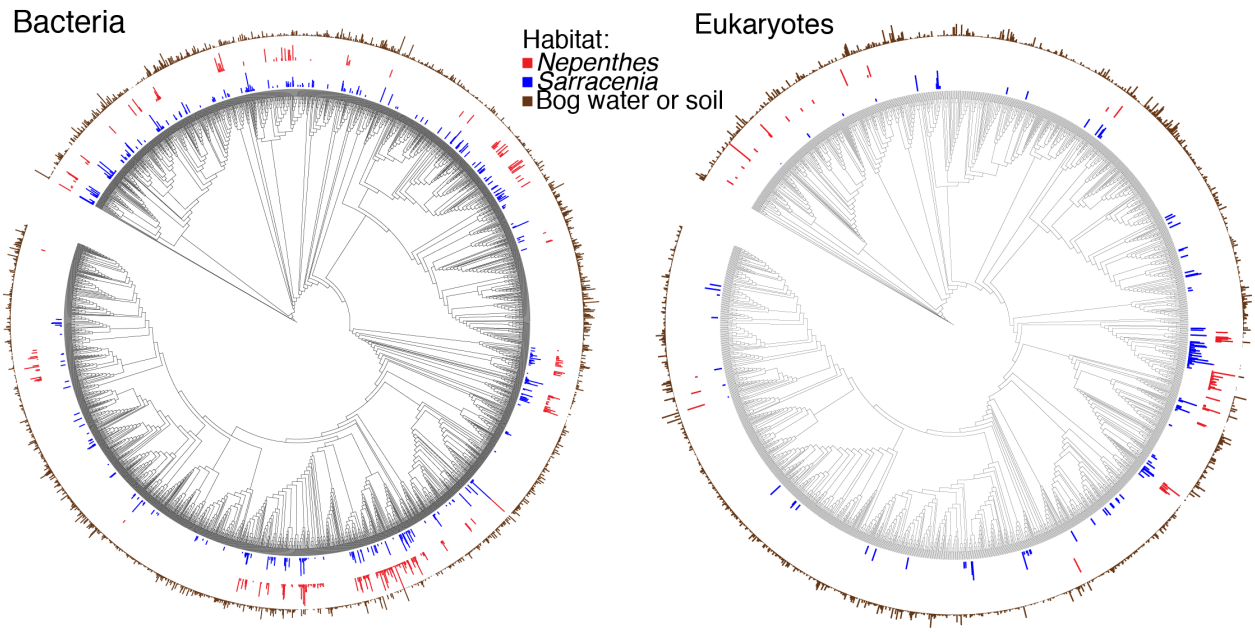


Figure 3.2: Phylogeny of bacterial and eukaryotic OTUs found in soil and bog samples (brown); and OTUs present in at least 10% of *Nepenthes* (red) and *Sarracenia* (blue) samples.

Within each genus, pitcher species, acidity, shape, and volume correlate with community composition

To understand more about which characteristics of pitchers shape internal community composition, we recorded species identity and pitcher growth form, and measured the pH and total volume of pitcher fluid for phytotelmata samples. In both *Sarracenia* and *Nepenthes* systems, pitcher species identity significantly influenced the internal communities (Figure 3.3, Supplementary Table 3.3). In bacteria, the effect of species was similar for *Sarracenia* (*envfit* $R^2 = 0.39$, $P < 0.001$; *adonis* $R^2 = 0.17$, $P < 0.001$) and *Nepenthes* (*envfit* $R^2 = 0.38$, $P < 0.001$; *adonis* $R^2 = 0.18$, $P < 0.001$); however, for eukaryotes, pitcher species explained more of the

observed variation in *Sarracenia* (*envfit* $R^2 = 0.42$, $P < 0.001$; *adonis* $R^2 = 0.30$, $P < 0.001$) than in *Nepenthes* (*envfit* $R^2 = 0.24$, $P < 0.001$; *adonis* $R^2 = 0.15$, $P < 0.001$) pitcher communities.

Pitcher species have different levels of acidity (which are more pronounced in *Nepenthes* than in *Sarracenia*) and different growth forms (more pronounced in *Sarracenia*). *Nepenthes* species can actively raise or lower the acidity of their pitchers by pumping protons into or out of pitcher fluid (An et al. 2001, Moran et al. 2010). The pH of pitcher fluid in some *Nepenthes* species quickly drops after opening, and then slowly increases again ((Hua and Li 2005) and personal observations). In our sampling of natural populations, we measured values below pH 4 in *N. rafflesiana*, *N. gracilis*, and *N. stenophylla*. The large pH gradient across the *Nepenthes* fluids in our samples was strongly correlated with bacterial community composition, and explained most of the observed variation (Figure 3.3A. *ordisurf* $R^2 = 0.74$, $P < 0.001$; *mantel* $r = 0.63$, $P < 0.001$). For *Nepenthes* eukaryotic community composition, pH was more weakly correlated and explained a smaller portion of the variation (*ordisurf* $R^2 = 0.21$, $P < 0.001$; *mantel* $r = 0.17$, $P < 0.001$).

Sarracenia species have distinct growth forms. For the species we sampled, *S. purpurea* and *S. rosea* have a shorter, more cylindrical shape, while *S. alata*, *S. flava*, *S. leucophylla* and *S. rubra* have a taller, more tapered shape. The taller, tapered *Sarracenia* pitchers have an aspect ratio below 0.2; while the shorter, more cylindrical Sarracenia have an aspect ratio above 0.2, as do the *Nepenthes* pitchers from this study. Shorter, wider pitchers tended to have a larger volume of fluid than the taller pitchers, thus we cannot separate the effects of pitcher volume and growth form. Both factors were correlated with *Sarracenia* bacterial (volume: *ordisurf* $R^2 = 0.31$, $P < 0.001$, *mantel* $r = 0.15$, $P = 0.007$; shape: *envfit* $R^2 = 0.31$, $P < 0.001$, *adonis* $R^2 = 0.08$, $P < 0.001$)

and eukaryotic community compositions (volume: *ordisurf* R^2= 0.17, P < 0.001, *mantel* r = 0.17, P = 0.002; shape: *envfit* R^2= 0.39, P < 0.001, *adonis* R^2= 0.21, P < 0.001).

For *Sarracenia* pitcher samples, we very rarely observed pH values below 4. Acidity of *Sarracenia* samples explained a much smaller portion of the variation in bacterial communities than those of *Nepenthes*, and was not a significant factor for eukaryotic communities (Supplementary Table 3.3). For *Nepenthes* pitchers, although volume varied more than in *Sarracenia*, the pitchers have similar growth forms despite being different sizes. Volume in *Nepenthes* samples explained a smaller portion of the variation in communities than those of *Sarracenia* (Supplementary Table 3.3), and we hypothesize that pitcher growth form has a stronger effect on community composition than fluid volume.

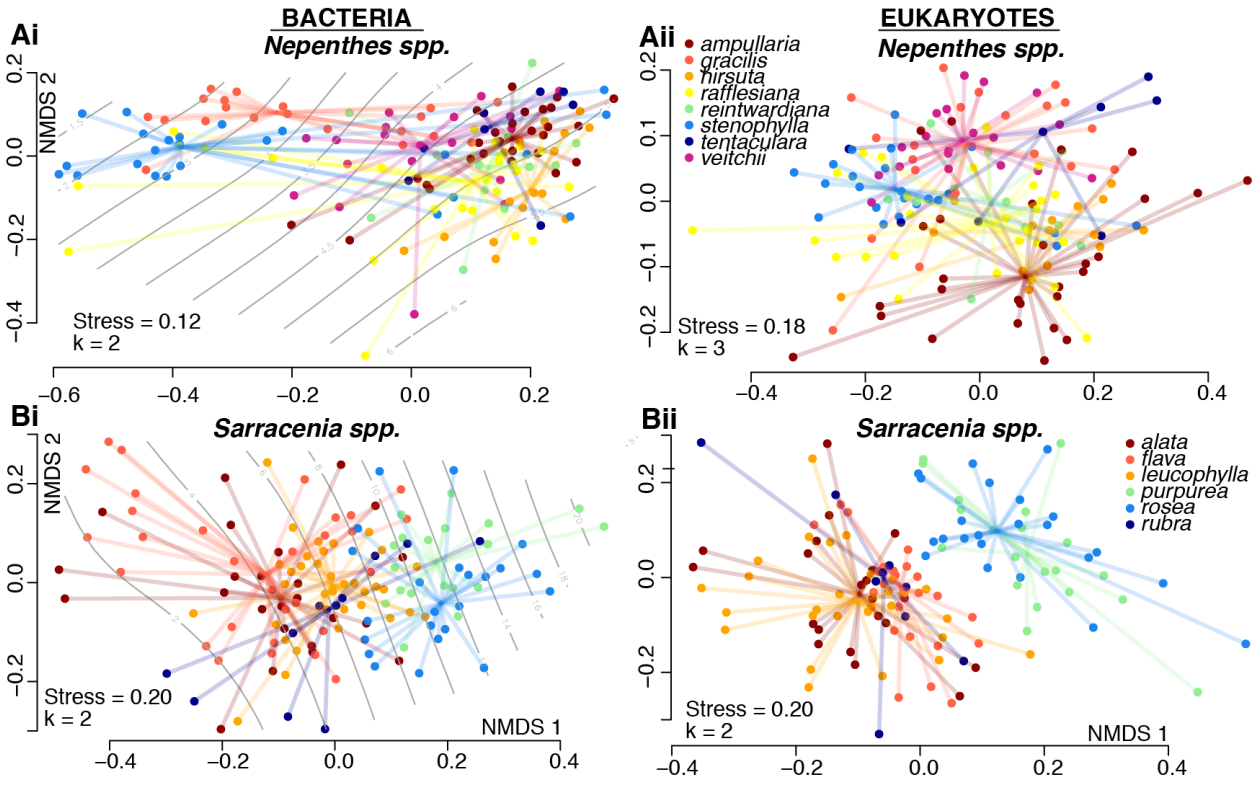


Figure 3.3: NMDS ordinations of pitcher samples, colored by host species. Ordisurf vectors with correlations greater than 0.3 are mapped into the ordinations: pH in (Ai) and volume in (Bi).

Relocated *Nepenthes* converge on *Sarracenia*-like communities

We conducted experiments where we relocated *Nepenthes* pitcher plants from Southeast Asia to a *Sarracenia* bog in North America, in order to test whether relocated *Nepenthes* pitchers would acquire communities more similar to those of their native region, or more similar to those of local *Sarracenia*. All *Nepenthes* placed into the *Sarracenia* habitat were maintained in pots with soil material purchased in the U.S., and removed after experiments were over, and we included potted *S. purpurea* as a control to explore whether growth in a pot influenced community assembly. Target pitchers were manually opened in the bog on Day 1 of each experiment. Experiments also included roughly pitcher-shaped, sterile glass tubes, either with or without sterilized insect material added as a nutrient control (Figure 3.4A). During the experiment, we noted the presence or absence of the pitcher plant mosquito, *Wyeomyia smithii*—an insect that completes its lifecycle only within *Sarracenia purpurea* pitchers (Figure 3.4B). *W. smithii* larvae regularly colonized their native *S. purpurea* pitchers (whether they were growing in the ground or in a pot). Surprisingly, they also colonized pitchers of *Nepenthes bicalcarata* and *N. ampullaria*, though in lower proportions (Figure 3.4B). However, the mosquitoes never colonized the more acidic *N. gracilis*, and *N. rafflesiana* species, or the experimental glass-tube pitchers.

To compare compositions of entire communities, we first re-plotted our beta-diversity results from natural *Nepenthes* versus *Sarracenia* pitchers, and found that community compositions were significantly different (Figure 3.4C). Bacterial communities in the two genera were more similar than eukaryotic communities, and correspondingly, genus explained less variation in bacterial than in eukaryotic community compositions (Bacteria: *envfit* $R^2 = 0.33$, $P <$

0.001, *adonis* $R^2 = 0.09$, $P < 0.001$; Eukaryota: *envfit* $R^2 = 0.56$, $P < 0.001$, *adonis* $R^2 = 0.14$, $P < 0.001$).

We then re-plotted community compositions of wild *Nepenthes* with our experimental results (Figure 3.4D). When *Nepenthes* communities assembled in a *Sarracenia* habitat, the communities of both bacteria and eukaryotes converged on similar compositions to those of the local *Sarracenia* and not wild *Nepenthes* (Figure 3.4D). The exception to this was in *Nepenthes* pitchers with pH below 4. The bacterial assemblages in highly-acidic pitchers (generally *N. gracilis* and *N. rafflesiana*) separated from other *Nepenthes* and *Sarracenia* pitchers in the same way that acidic pitcher bacterial communities shifted in natural *Nepenthes* populations (Figures 3.3A and 3.5D). Acidity explained most of the variation in bacterial community composition from the experiments (pH: *ordisurf* $R^2 = 0.67$, $P < 0.001$, *mantel* $r = 0.50$, $P < 0.001$) and was also a significant predictor of eukaryotic community composition (pH: *ordisurf* $R^2 = 0.23$, $P < 0.001$, *mantel* $r = 0.12$, $P < 0.001$). Region—whether the pitchers were in Harvard Forest, Singapore or Malaysia—explained a surprisingly small portion of bacterial variation (*envfit* $R^2 = 0.19$, $P < 0.001$, *adonis* $R^2 = 0.07$, $P < 0.001$), but a larger portion of the eukaryotic variation (*envfit* $R^2 = 0.41$, $P < 0.001$, *adonis* $R^2 = 0.12$, $P < 0.001$).

Experimental bacterial and eukaryotic communities in pitchers were different from bog water communities (not shown), but partially clustered with the organisms colonizing the glass tube pitchers (Figure 3.4D). This results suggests a sterile pitcher-shaped form is sufficient for acquiring a pitcher plant-like microbiome; however, the glass tube communities were only a subset of the full variety of natural pitcher plant communities (Figure 3.4D).

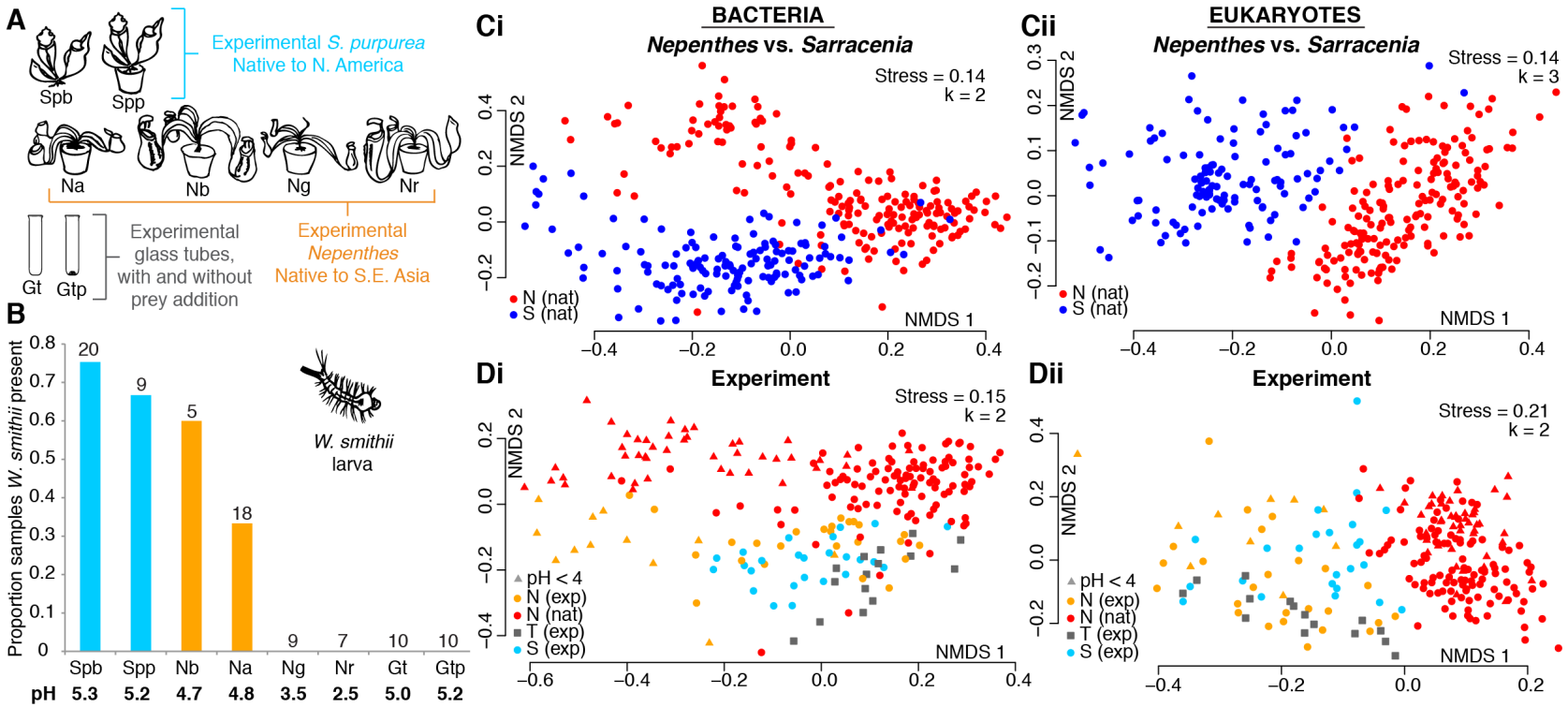


Figure 3.4: A) Experimental treatments. B) *Wyeomyia smithii* (pitcher plant mosquito) larvae colonized their native *S. purpurea* hosts, as well as foreign *Nepenthes* species with average pH > 4, but not pitcher-shaped glass tubes. Numbers of samples are listed above each category, and average pH values are listed below. C) Natural *Nepenthes* and *Sarracenia* pitchers house different organisms (Ci and Cii). D) Experimentally-relocated *Nepenthes* converge on *Sarracenia*-like communities and differ from those of natural *Nepenthes* (Di and Dii), except for bacterial communities in samples where pH < 4 (Di). Glass tubes with a pitcher-like form assemble communities that are similar to those of experimental pitchers (Di and Dii).

Nepenthes and _Sarracenia_ have different total functional potential; both have high relative abundances of chitinases

We generated metagenomes of 24 pitcher samples (16 *Nepenthes* and 8 *Sarracenia*) to investigate functional potential within pitcher microbiomes. When compared with other published metagenomes for soil, lake, and phyllosphere samples (Supplementary Table 3.4), pitcher plant community metagenomes were most similar to other phyllosphere communities in overall functional potential as measured by Kegg Orthology (KO) groups using the Bray-Curtis dissimilarity metric in an NMDS plot (Figure 3.5A). Among *Nepenthes* and *Sarracenia* metagenomes, KO gene families in *Sarracenia purpurea* clustered close to *Nepenthes ampullaria*, *N. gracilis* and *N. reintwardiana*. Other *Sarracenia* (those with a tapered shape) and other *Nepenthes* (those with more acidic fluid) were more different in terms of functional potential. We wanted to test functional similarity of *Nepenthes* and *Sarracenia* communities using a function likely to impact the interaction between host and microbiome, so we investigated abundances of chitinases (K01183, GH families 18 and 19). As chitin is the main component of insect exoskeletons, we expected pitcher plant microbiomes would have the potential to digest it. *Nepenthes* and *Sarracenia* microbiome metagenomes had high relative abundances of chitinase genes as compared to other habitats (Figure 3.5B). Combined pitcher plant relative chitinase abundances where higher than those in other combined metagenomes (Mann-Whitney U Test, $P = 0.0197$).

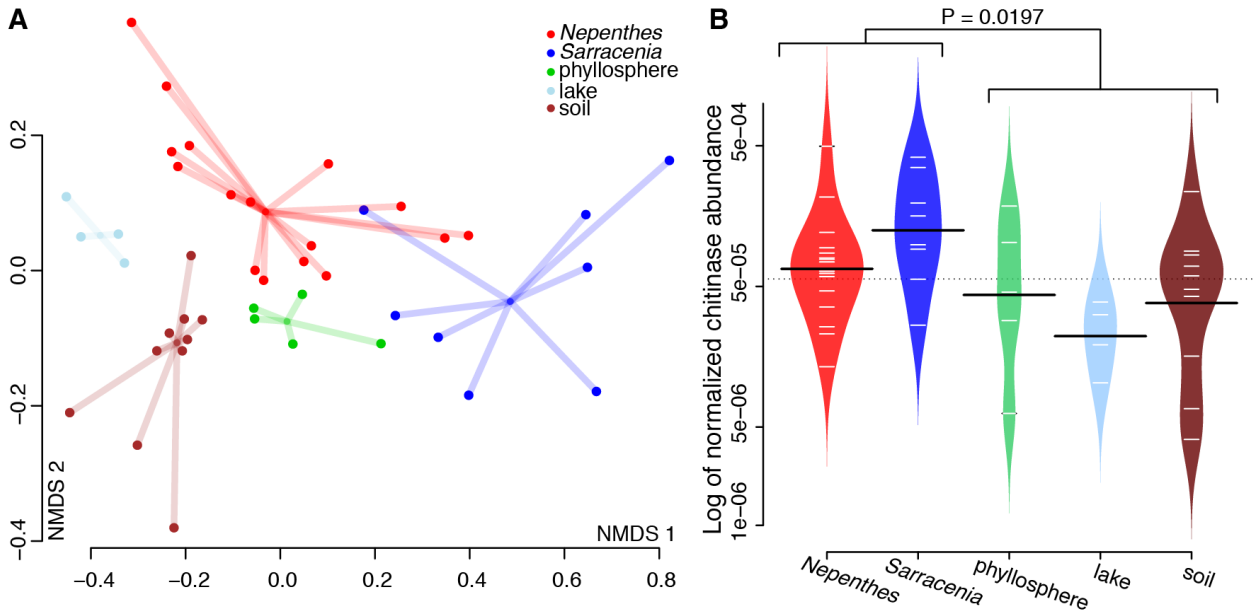


Figure 3.5: A) NMDS plot of functional gene families comparing pitcher plant and environmental metagenomes. B) Relative abundance of chitinase gene family in metagenomes.

Discussion

For over a century, ecologists have been exploring which factors influence the organization of plant communities (e.g., Clements 1916, Gleason 1926), and, more recently, microbial communities (e.g., Lindström and Langenheder 2012). However, due to the complexity of microbial communities, it is very difficult to pinpoint which characteristics structure them. One approach to understanding how microbial communities assemble is to examine host-associated systems. Microbes interact with virtually all plants and animals (Van Der Heijden et al. 2008, McFall-Ngai et al. 2013), and convergent plant or animal morphologies may influence the structure and function of associated microbial communities (Bittleston et al. 2016). The physical structure and local chemistry of host habitats can regulate how communities form within them. Furthermore, convergent evolution in the form and function of a host can be used as a tool to uncover the factors influencing community assembly, and the ways in which the influence manifests.

We investigated if (and in which ways) the convergent morphology and function of pitcher plant's modified leaves lead to convergence of associated communities. In their natural environments, we found that bacterial and eukaryotic communities in both *Nepenthes* and *Sarracenia* pitchers have lower taxonomic richness and evenness than in surrounding soil or bog water, suggesting that pitcher habitats favor a subset of available species. Although the communities within pitchers of the two genera contained different organisms, the organisms tended to be from similar phylogenetic clades, particularly for bacteria. These phylogenetically related organisms might have corresponding functions within *Nepenthes* and *Sarracenia* pitcher micro-ecosystems. In an experimental context, when placed in the same environment, the community compositions of *Nepenthes* and *Sarracenia* pitchers converged, as long as pitcher-

fluid acidity levels were comparable. The cup or tube-like shape of the pitcher plant's modified leaves had a strong influence on community assembly, and other, unmeasured aspects of the pitcher habitat caused organisms adapted to *Sarracenia purpurea* pitchers to also colonize *Nepenthes* pitchers, but not artificial pitchers.

The pitcher shape, and the way it captures rainwater, appears to regulate fundamental aspects of pitcher community diversity. When *Nepenthes* and *Sarracenia* pitchers were set in a common environment, community compositions converged in pitcher fluids with comparable pH levels. Surprisingly, pitcher-shaped glass tubes also converged on similar community composition (Figure 3.4D). The tubes mimic a pitcher shape and trap occasional prey, but do not share color, texture, or chemistry with living pitchers. In Southeast Asia, we also found that plastic tubes collected communities as similar—sometimes even more similar—to *Nepenthes* pitchers as water held in fallen leaves (Supplementary Figure 3.1). However, although pitcher-shaped tubes overlapped with living pitchers in terms of community composition, in our experiment some *Sarracenia* and *Nepenthes* pitcher communities were still more similar to each other than to tube communities (Figure 3.4D). Moreover, the pitcher plant mosquito, *Wyeomyia smithii*, which normally breeds only within *S. purpurea* pitchers, colonized pitchers of *N. bicalcarata* and *N. ampullaria* but not pitcher-shaped glass tubes (Figure 3.4B). Thus, unmeasured characteristics of plant-formed pitchers also regulate the assembling communities, although to a lesser extent than general pitcher morphology.

To further understand factors influencing pitcher microcosms, we examined bacterial and eukaryotic communities in different plant species within each genus of pitcher plants. Communities living within pitcher ecosystems are clearly influenced by characteristics of their host species, and these effects explain more of the variation in community structure than does

geographic region. In our sampling of natural *Sarracenia* and *Nepenthes* populations in North America and Southeast Asia, we found significant differences in bacterial beta diversity among different pitcher plant species. Previous studies have demonstrated that pH is a strong predictor of bacterial community composition in soils (Fierer and Jackson 2006, Lauber et al. 2009, Rousk et al. 2010), and we see the same effect of pH on *Nepenthes* pitcher communities. *Nepenthes* species can actively acidify their pitcher fluid (An et al. 2001, Moran et al. 2010), and in a few species pH has been observed to first decrease, and then increase over time ((Hua and Li 2005) and personal observations). In *N. gracilis*, *N. rafflesiana* and *N. stenophylla* species, pitcher community compositions in acidic pitcher fluids were very different from those in less acidic fluids of the same species. These results indicate that acidity has a stronger effect on bacterial community composition than does host species or geographic region alone. However, among samples with similar pH levels, we still found bacterial community composition clustering by host species. Acidity was a weaker predictor of eukaryotic community composition, which also corroborates previous findings in soil communities (Rousk et al. 2010). Even when pH levels are similar, pitcher plant species identity influences eukaryotic community composition, likely due to characteristics of host species not measured in this study.

For *Sarracenia* pitchers, differences among community compositions were correlated with the volume and general shape of pitchers, as well as collection site. *Sarracenia purpurea* and *S. rosea* pitchers are more squat and cylindrical and tend to contain more fluid, while *S. alata*, *S. flava*, *S. leucophylla*, and *S. rubra* have tall, tapered pitchers that tend to contain smaller volumes of fluid. Bacterial community compositions of *Sarracenia* pitchers were most strongly influenced by pitcher species and total volume of the pitchers, while eukaryotic community compositions clustered by pitcher species and general pitcher shape. Collection site also

significantly influenced *Sarracenia* communities; however, the effect was weaker when we controlled for variation caused by different species growing at different sites (Supplementary Table 3.3).

Metagenomic functional analyses revealed high variability among pitcher communities, and similarities between *Nepenthes* and *Sarracenia* systems. When comparing beta diversity of gene families (determined by Kegg Orthology groupings) among our samples and published metagenomes from lakes, soils, and phyllospheres, pitcher functional potential was most similar to phyllosphere metagenomes (Figure 3.5A). A recent study of *Nepenthes* bacterial communities also found similarities among pitcher and other published phyllosphere communities (Takeuchi et al. 2015). When comparing *Nepenthes* and *Sarracenia* microbiome functional potential, samples from species with more similar shape and chemistry had more similar overall functional gene families. *Nepenthes* species that were more acidic and *Sarracenia* species with a fluted shape were less similar.

Because trapping and digesting arthropod prey is a central feature of pitcher plant ecology, and because arthropod exoskeletons primarily consist of chitin, we expected to see high levels of chitinases in pitcher plant microbiomes. Chitinase gene family abundances are higher on average in pitcher plant metagenomes than those of other environmental metagenomes. *Nepenthes* plants produce some of their own chitinase enzymes (e.g. see Eilenberg et al. 2006), but we would still expect microbes with chitinases to benefit in this system. As far as we know, *Sarracenia* plants do not produce their own chitinases (Adlassnig et al. 2011), and in fact we see the highest levels of chitinase genes in *Sarracenia* phytotelmata metagenomes (Figure 3.5B). It may be particularly beneficial for both the plants and their microbial inhabitants if the microbes can excrete chitinase enzymes, making more nitrogen and carbon available to the common pool.

Organisms from the same phylogenetic groups that are repeatedly found living in pitcher systems—for example, bacteria in the families Rhizobiaceae and Chitinophagaceae, dipteran insects, aquatic mites, and bdelloid rotifers—may fulfill similar functional roles within pitcher ecosystems. Aquatic mites and midge larvae act as detritivores, shredding prey items and contributing to nutrient cycling (Beaver 1983, Kitching 2000). Mosquito larvae feed on protozoa and bacteria, and can regulate microbial populations (Peterson et al. 2008). Other mosquitoes and dipteran insects act as top predators, feeding on smaller living insects. Bacteria in the Rhizobiaceae family can fix atmospheric nitrogen (Fischer 1994) and may add nutrients to the aquatic pools, while some members of Chitinophagaceae can digest chitin (Rosenberg 2014).

The convergent interactions of pitcher plants and their microbiomes might be used to predict microbial biodiversity in other systems. A promising future direction will be to test if the results from this study can predict microbial biodiversity in pitchers of *Cephalotus follicularis*, the sole species of a third, independently-evolved pitcher plant family. In addition, future studies can test if the factors influencing pitcher communities have the same effects in other contained, aquatic ecosystems. For example, we would expect to see lower levels of species richness in other small, ephemeral systems such as bamboo internodes and tree holes. Furthermore, we predict that pH levels below 4 will affect bacterial community compositions more than differences in volume or geographic region. In contexts where the acidity levels are more similar, we would expect volume and other aspects of shape to more strongly affect the communities.

Convergent interactions among microbes and macro-organisms have been found in other systems as well. Bacterial communities associated with fungus-growing ants, beetles, and termites in different regions of the world have dominant communities members from the same genera, with convergent functional potential (Aylward et al. 2014). Microbial symbionts of

sponges are functionally equivalent across phylogenetically divergent hosts (Fan et al. 2012). Gut bacteria of myrmecophagous animals from different taxonomic orders have convergent community composition, driven by a similar ant diet (Delsuc et al. 2014). Convergent interactions can be used to better understand the selective pressures structuring microbial biodiversity, and can provide predictions to be tested across systems. Instead of defining and counting species, we can explore compositional and functional similarities of whole communities in order to accelerate the mapping of global microbial diversity and the development of conservation strategies.

Chapter 3 References:

Abubucker, S., N. Segata, J. Goll, A. M. Schubert, J. Izard, B. L. Cantarel, B. Rodriguez-Mueller, J. Zucker, M. Thiagarajan, B. Henrissat, O. White, S. T. Kelley, B. Methé, P. D. Schloss, D. Gevers, M. Mitreva, and C. Huttenhower. 2012. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS computational biology 8:e1002358.

Adlassnig, W., M. Peroutka, and T. Lendl. 2011. Traps of carnivorous pitcher plants as a habitat: composition of the fluid, biodiversity and mutualistic activities. Annals of botany 107:181–194.

Albert, V. A., S. E. Williams, and M. W. Chase. 1992. Carnivorous plants: phylogeny and structural evolution. Science 257:1491–1495.

Amaral-Zettler, L. A., E. A. McCliment, H. W. Ducklow, and S. M. Huse. 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PloS one 4:e6372.

An, C.-I., E.-I. Fukusaki, and A. Kobayashi. 2001. Plasma-membrane H+-ATPases are expressed in pitchers of the carnivorous plant *Nepenthes alata* Blanco. Planta 212:547–555.

Aylward, F. O., G. Suen, P. H. W. Biedermann, A. S. Adams, J. J. Scott, S. A. Malfatti, T. Glavina del Rio, S. G. Tringe, M. Poulsen, K. F. Raffa, K. D. Klepzig, and C. R. Currie. 2014. Convergent bacterial microbiotas in the fungal agricultural systems of insects. mBio 5:e02077.

Bates, S. T., J. C. Clemente, G. E. Flores, W. A. Walters, L. W. Parfrey, R. Knight, and N. Fierer. 2012. Global biogeography of highly diverse protistan communities in soil. The ISME Journal 7:652–659.

Beaver, R. A. 1983. The communities living in *Nepenthes* pitcher plants: fauna and food webs. Pages 129–159 *in* J. H. Frank and L. P. Lounibos, editors. Phytotelmata: Plants as hosts for aquatic insect communities. Plexus Pub., New Jersey.

Bittleston, L. S., C. Baker, L. B. Strominger, and A. Pringle. 2015. Metabarcoding as a tool for investigating arthropod diversity in *Nepenthes* pitcher plants. Austral Ecology.

Bittleston, L. S., N. E. Pierce, A. M. Ellison, and A. Pringle. 2016. Convergence in multispecies Interactions. Trends in ecology & evolution 31:269–280.

Bokulich, N. A., S. Subramanian, J. J. Faith, D. Gevers, J. I. Gordon, R. Knight, D. A. Mills, and J. G. Caporaso. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods 10:57–59.

Bradshaw, W. E., and R. A. Creelman. 1984. Mutualism between the carnivorous purple

pitcher plant and its inhabitants. American Midland Naturalist 112:294–304.

Buch, F., M. Rott, S. Rottloff, C. Paetz, I. Hilke, M. Raessler, and A. Mithöfer. 2013. Secreted pitfall-trap fluid of carnivorous *Nepenthes* plants is unsuitable for microbial growth. Annals of botany 111:375–383.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, and M. Bauer. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal 6:1621–1624.

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

Chan, X. Y., K. W. Hong, W. F. Yin, and K. G. Chan. 2016. Microbiome and biocatalytic bacteria in monkey cup (*Nepenthes* pitcher) digestive fluid. Scientific reports.

Chapin, C. T., and J. Pastor. 1995. Nutrient limitations in the northern pitcher plant *Sarracenia purpurea*. Canadian Journal of Botany 73:728–734.

Chou, L. Y., C. M. Clarke, and G. A. Dykes. 2014. Bacterial communities associated with the pitcher fluids of three *Nepenthes* (Nepenthaceae) pitcher plant species growing in the wild. Archives of microbiology.

Clarke, C. M., and R. L. Kitching. 1993. The metazoan food webs from six Bornean *Nepenthes* species. Ecological Entomology 18:7–16.

Clements, F. E. 1916. Plant Succession: An Analysis of the Development of Vegetation. Carnegie Institution of Washington, Washington D.C.

Davison, J., M. Moora, M. Opik, A. Adholeya, L. Ainsaar, A. Ba, S. Burla, A. G. Diedhiou, I. Hiiesalu, T. Jairus, N. C. Johnson, A. Kane, K. Koorem, M. Kochar, C. Ndiaye, M. Partel, U. Reier, U. Saks, R. Singh, M. Vasar, and M. Zobel. 2015. Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. Science 349:970–973.

Delsuc, F., J. L. Metcalf, L. Wegener Parfrey, S. J. Song, A. González, and R. Knight. 2014. Convergence of gut microbiomes in myrmecophagous mammals. Molecular Ecology 23:1301–1317.

Dirzo, R., H. S. Young, M. Galetti, G. Ceballos, N. J. B. Isaac, and B. Collen. 2014. Defaunation in the Anthropocene. Science 345:401–406.

Eilenberg, H., S. Pnini-Cohen, S. Schuster, A. Movtchan, and A. Zilberstein. 2006. Isolation and characterization of chitinase genes from pitchers of the carnivorous plant

Nepenthes khasiana. Journal of Experimental Botany 57:2775–2784.

Ellison, A. M. 2006. Nutrient limitation and stoichiometry of carnivorous plants. Plant biology (Stuttgart, Germany) 8:740–747.

Fan, L., D. Reynolds, M. Liu, M. Stark, S. Kjelleberg, N. S. Webster, and T. Thomas. 2012. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. Proceedings of the National Academy of Sciences of the United States of America 109:E1878–87.

Fierer, N., and R. B. Jackson. 2006. The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences of the United States of America 103:626–631.

Fischer, H. M. 1994. Genetic regulation of nitrogen fixation in rhizobia. Microbiological reviews 58:352–386.

Gibbons, S. M., J. G. Caporaso, M. Pirrung, D. Field, R. Knight, and J. A. Gilbert. 2013. Evidence for a persistent microbial seed bank throughout the global ocean. Proceedings of the National Academy of Sciences of the United States of America 110:4651–4655.

Glass, E. M., J. Wilkening, A. Wilke, D. Antonopoulos, and F. Meyer. 2010. Using the Metagenomics RAST Server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harbor Protocols 2010:pdb.prot5368–pdb.prot5368.

Gleason, H. A. 1926. The individualistic concept of the plant association. Bulletin of the Torrey Botanical Club 53:7–26.

Heard, S. B. 1998. Capture rates of invertebrate prey by the pitcher plant, *Sarracenia purpurea* L. American Midland Naturalist:79–89.

Hua, Y., and H. Li. 2005. Food web and fluid in pitchers of *Nepenthes mirabilis* in Zhuhai, China. Acta Botanica Gallica 152:165–175.

Jetz, W., D. S. Wilcove, and A. P. Dobson. 2007. Projected impacts of climate and land-use change on the global diversity of birds. PLoS biology 5:e157.

Juniper, B. B. E., R. J. Robins, and D. M. Joel. 1989. The carnivorous plants. Academic Press.

Kampstra, P. 2008. Beanplot: A boxplot alternative for visual comparison of distributions. Journal of statistical software.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics (Oxford, England) 26:1463–1464.

Kitching, R. L. 2000. Food webs and container habitats. Cambridge University Press.

Kneitel, J. M., and T. E. Miller. 2002. Resource and top-predator regulation in the pitcher plant (*Sarracenia purpurea*) inquiline community. Ecology 83:680–688.

Koopman, M. M., D. M. Fuselier, S. Hird, and B. C. Carstens. 2010. The carnivorous pale pitcher plant harbors diverse, distinct, and time-dependent bacterial communities. Applied and environmental microbiology 76:1851–1860.

Krieger, J. R., and P. S. Kourtev. 2012. Bacterial diversity in three distinct sub-habitats within the pitchers of the northern pitcher plant, *Sarracenia purpurea*. FEMS Microbiology Ecology 79:555–567.

Kurup, R., A. J. Johnson, S. Sankar, A. A. Hussain, C. S. Kumar, and B. Sabulal. 2013. Fluorescent prey traps in carnivorous plants. Plant Biology 15:611–615.

Lauber, C. L., M. Hamady, R. Knight, and N. Fierer. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Applied and environmental microbiology 75:5111–5120.

Letunic, I., and P. Bork. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic acids research 39:W475–8.

Lindström, E. S., and S. Langenheder. 2012. Local and regional factors influencing bacterial community assembly. Environmental microbiology reports 4:1–9.

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

McFall-Ngai, M., M. G. Hadfield, T. C. G. Bosch, H. V. Carey, T. Domazet-Lošo, A. E. Douglas, N. Dubilier, G. Eberl, T. Fukami, S. F. Gilbert, U. Hentschel, N. King, S. Kjelleberg, A. H. Knoll, N. Kremer, S. K. Mazmanian, J. L. Metcalf, K. Nealson, N. E. Pierce, J. F. Rawls, A. Reid, E. G. Ruby, M. Rumpho, J. G. Sanders, D. Tautz, and J. J. Wernegreen. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proceedings of the National Academy of Sciences of the United States of America 110:3229–3236.

McMurdie, P. J., and S. Holmes. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one 8:e61217.

Moran, J. A., B. J. Hawkins, B. E. Gowen, and S. L. Robbins. 2010. Ion fluxes across the pitcher walls of three Bornean *Nepenthes* pitcher plant species: flux rates and gland distribution patterns reflect nitrogen sequestration strategies. Journal of Experimental Botany 61:1365–1374.

Nelson, M. C., H. G. Morrison, J. Benjamino, S. L. Grim, and J. Graf. 2014. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. PloS one 9:e94249.

Nemergut, D. R., E. K. Costello, M. Hamady, C. Lozupone, L. Jiang, S. K. Schmidt, N. Fierer, A.

R. Townsend, C. C. Cleveland, L. Stanish, and R. Knight. 2011. Global patterns in the biogeography of bacterial taxa. Environmental microbiology … 13:135–144.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2013, December. vegan: Community Ecology Package.

Osunkoya, O. O., S. D. Daud, and F. L. Wimmer. 2008. Longevity, lignin content and construction cost of the assimilatory organs of *Nepenthes* species. Annals of botany 102:845–853.

Peterson, C. N., S. Day, B. E. Wolfe, A. M. Ellison, R. Kolter, and A. Pringle. 2008. A keystone predator controls bacterial diversity in the pitcher-plant (*Sarracenia purpurea*) microecosystem. Environmental microbiology … 10:2257–2266.

Rosenberg, E. 2014. The Family Chitinophagaceae. Pages 493–495 *in* The Prokaryotes. Springer Berlin Heidelberg, Berlin, Heidelberg.

Rousk, J., E. Bååth, P. C. Brookes, C. L. Lauber, C. Lozupone, J. G. Caporaso, R. Knight, and N. Fierer. 2010. Soil bacterial and fungal communities across a pH gradient in an arable soil. The ISME Journal 4:1340–1351.

Sambrook, J., and D. W. Russell. 2001. Molecular cloning: a laboratory manual. Third edition. Coldspring-Harbour Laboratory Press, UK.

Schulze, W., E. D. Schulze, J. S. Pate, and A. N. Gillison. 1997. The nitrogen supply from soils and insects during growth of the pitcher plants *Nepenthes mirabilis*, *Cephalotus follicularis* and *Darlingtonia californica*. Oecologia 112:464–471.

Sickel, W., T. U. Grafe, I. Meuche, and I. Steffan-Dewenter. 2016. Bacterial diversity and community structure in two Bornean *Nepenthes* species with differences in nitrogen acquisition strategies. Microbial ecology.

Takeuchi, Y., S. Chaffron, and M. M. Salcher. 2015. Bacterial diversity and composition in the fluid of pitcher plants of the genus *Nepenthes*. Systematic and Applied Microbiology:330–339.

Van Der Heijden, M. G., R. D. Bardgett, and N. M. van Straalen. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecology Letters 11:296–310.

Appendix A: Dissecting host-associated communities with DNA barcodes

Manuscript accepted and in press at Philosophical Transactions of the Royal Society B.

Authors: Christopher CM Baker[1,2], Leonora S Bittleston[1], Jon G Sanders[1,3], Naomi E Pierce[1,*]

Addresses:

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge MA 02138, USA.

[2]Present address: Department of Ecology and Evolutionary Biology, Princeton University, Princeton NJ 08544, USA.

[3]Present address: Department of Pediatrics, University of California San Diego, San Diego CA 92093, USA.

ORCID IDs: CCMB: 0000-0002-2675-1078; LSB: 0000-0003-4007-5405; JGS: 0000-0001-6077-4014

***Author for correspondence**: Naomi E Pierce, email: npierce@oeb.harvard.edu

Keywords: *Nepenthes*, *Vachellia drepanolobium*, *Cephalotes*, DNA barcoding, metabarcoding, species interactions

Summary

DNA barcoding and metabarcoding methods have been invaluable in the study of interactions between host organisms and their symbiotic communities. Barcodes can help identify individual symbionts that are difficult to distinguish using morphological characters, and provide a way to classify undescribed species. Entire symbiont communities can be characterized rapidly using barcoding and especially metabarcoding methods, which is often crucial for isolating ecological signal from the substantial variation among individual hosts. Furthermore, barcodes allow the evolutionary histories of symbionts and their hosts to be assessed simultaneously and in reference to one another. Here we describe three projects illustrating the utility of barcodes for studying symbiotic interactions: first, we consider communities of arthropods found in the ant-occupied domatia of the East African ant-plant *Vachellia (Acacia) drepanolobium*; second, we examine communities of arthropod and protozoan inquilines in three species of *Nepenthes* pitcher plant in South East Asia; third, we investigate communities of gut bacteria of South American ants in the genus *Cephalotes*. Advances in sequencing and computation, and greater database connectivity, will continue to expand the utility of barcoding methods for the study of species interactions, especially if barcoding can be approached flexibly by making use of alternative genetic loci, metagenomes and whole genome data.

Introduction

In many species interactions, a host organism associates with a community of symbionts. Bacteria and protozoa in guts of lower termites, for example, help their hosts obtain nutrition from digestion-resistant foods [1]. Some 300 species of insects and mites have been found accompanying colonies of *Eciton burchellii* army ants and are known to depend at least in part

on the ants [2]. Lichens, themselves symbioses of fungi and algae or cyanobacteria, host distinctive communities of bacteria on their surfaces, including lineages known almost exclusively from lichens [3]. These kinds of interactions are distinguished from simpler host-symbiont relationships by the potential for interactions among symbionts, and from studies of communities in abiotic contexts by the role of selection and phylogeny in shaping host interactions with symbionts. While species associations such as these have long been studied with a variety of approaches, DNA barcoding methods have in recent times become a useful addition to researchers' toolkits.

Barcoding can help classify symbiont taxa that would otherwise be difficult to identify. For many symbionts, morphological characters are inconspicuous or insufficient for identification, and for these organisms DNA identification may be helpful. Insect juveniles, such as those associated with ant colonies, often have few good identifying characters; bacteria and fungi likewise can be hard to identify. In these cases, DNA identification may help reduce the time and effort for identification, or may help identify cryptic species (e.g. [4-6]). Even where species are undescribed or are not included in sequences databases, similarity-based clustering of DNA barcodes allow organisms to be placed into groups that may be treated like species; such groups are often referred to as 'operational taxonomic units' or OTUs [7]. Since only small quantities of DNA are required, barcoding methods in general have broad application for species identification – they do not necessarily require intact specimens and can therefore be used with samples ranging from soil for biodiversity assessment [8], to fecal samples for diet analysis [9, 10], and even previously-parasitized leaf samples for the identification of emerged leaf miners and their parasitoids [11].

Using DNA barcodes can also provide insight into organization at the level of the whole community, by facilitating the rapid profiling of entire symbiont communities. Symbiont taxa often vary considerably among individual hosts, as well as between different host taxa or habitats, and parsing this variation requires analysis of the symbiont communities associated with many individual hosts. Of course this is not specific to mutualistic symbionts, and indeed barcoding has been used to good effect across a wide range of species interactions, such as assessing variation in parasitoid communities [12]. Furthermore, some patterns, such as interactions among the symbiont taxa themselves, may only be visible if the whole symbiont community is considered [13]. Community-level analysis has been especially pertinent to microbial symbioses, such as gut bacterial communities. In these cases, the combination of DNA barcoding with high-throughput sequencing technologies has facilitated the taxonomic profiling of complex communities through the simultaneous sequencing of many thousands of DNA barcodes from each sample, often termed ‘metabarcoding’.

DNA barcodes also permit the analysis of species interactions on evolutionary timescales. DNA barcodes are not just arbitrary species labels but, like any other part of the genome, contain the signature of their evolutionary past: recently diverged taxa tend to have more similar sequences than distantly related taxa. Using barcode data to compare evolutionary relationships among host taxa with those among symbiont taxa potentially provides a way to detect relevant patterns in those evolutionary histories, such as the co-diversification between hosts and symbionts.

In this paper, we review three DNA barcode-based studies we have performed that demonstrate the broad scope for using DNA barcodes to study species interactions. First, we outline our study of arthropods residing in the hollow, swollen thorns of the African ant-plant

Vachellia (Acacia) drepanolobium based on COI barcodes. Second, we describe our use of 18S metabarcoding to identify arthropods and arthropod-associated protozoa in *Nepenthes* pitcher plants. Third, we detail our use of 16S metabarcoding to explore codiversification of gut bacterial communities with their *Cephalotes* ant hosts. Our studies serve to illustrate the scope and flexibility of barcodes as analytical tools in the study of species interactions.

Myrmecophile communities in *Vachellia drepanolobium* ant plants

DNA barcoding has proven valuable for examining communities of arthropods residing in domatia of the ant-plant *Vachellia (Acacia) drepanolobium*.

V. drepanolobium is widespread throughout the East African tropics, often forming large mono-dominant stands in savannas with hardpan grey soil or poorly-drained black cotton soil (Figure A.1a) [14]. *V. drepanolobium* is covered with hollow swollen-thorn domatia (Figure A.1b) that, at least on larger trees, are usually occupied by ants [14]. Three ant species nest obligately in these domatia: *Crematogaster mimosae*, *C. nigriceps, and Tetraponera penzigi* [15]. A fourth species, *C. sjostedti*, also associates with *V. drepanolobium* trees but more commonly nests in trunk cavities created by cerambycid beetles or in the ground around the tree bases [16]. Each tree is normally occupied by a single ant species, but different trees, even within meters of one another, may be occupied by different species [16].

The obligate domatium-dwelling ants engage in a classic protection mutualism [17] with their hosts. In exchange for housing, as well as food in the form of extrafloral nectar, the ants protect their host plant from mammalian herbivores such as elephants, giraffe and antelope [18-20]. The ants vary, however, in the quality of their defense [15, 21]. Among the three domatium-dwelling ants, the aggressive *C. mimosae* provides better defense than *C. nigriceps*, while

T. penzigi does little to deter browsers [22]. And the ants impose other costs on their hosts: *C. nigriceps* prunes the plant's axillary buds, shaping growth and temporarily preventing flowering, while *T. penzigi* prunes the extrafloral nectaries, perhaps to reduce the risk of invasion by another ant colony [23, 24].

The ants' effects are also evident in the diverse assemblage of other organisms found on the host plant. A 2012 study of insects in the tree canopy, using a morphospecies approach, found that canopy communities on trees occupied by *C. mimosae* and *C. nigriceps* were distinct from those on trees occupied by *C. sjostedti* and *T. penzigi* [25]. Other studies of specific tree inhabitants also describe preferential associations with ant species (e.g. [15, 26]). Scale insects, for example, are found with *C. mimosae* and *C. sjostedti* [14], while neither *C. nigriceps* nor *T. penzigi* is typically found with scales. The lycaenid *Anthene usamba* specializes on trees occupied by *C. mimosae* [27]. The braconid wasp *Trigastrotheca laikipiensis* is a brood parasite of claustral colonies of *C. mimosae* and *C. nigriceps*, but is rarely found with *T. penzigi* [28], and *Acacidiplosis* gall midge parasites are found more frequently with *C. mimosae* ants than with *C. nigriceps* ants [29].

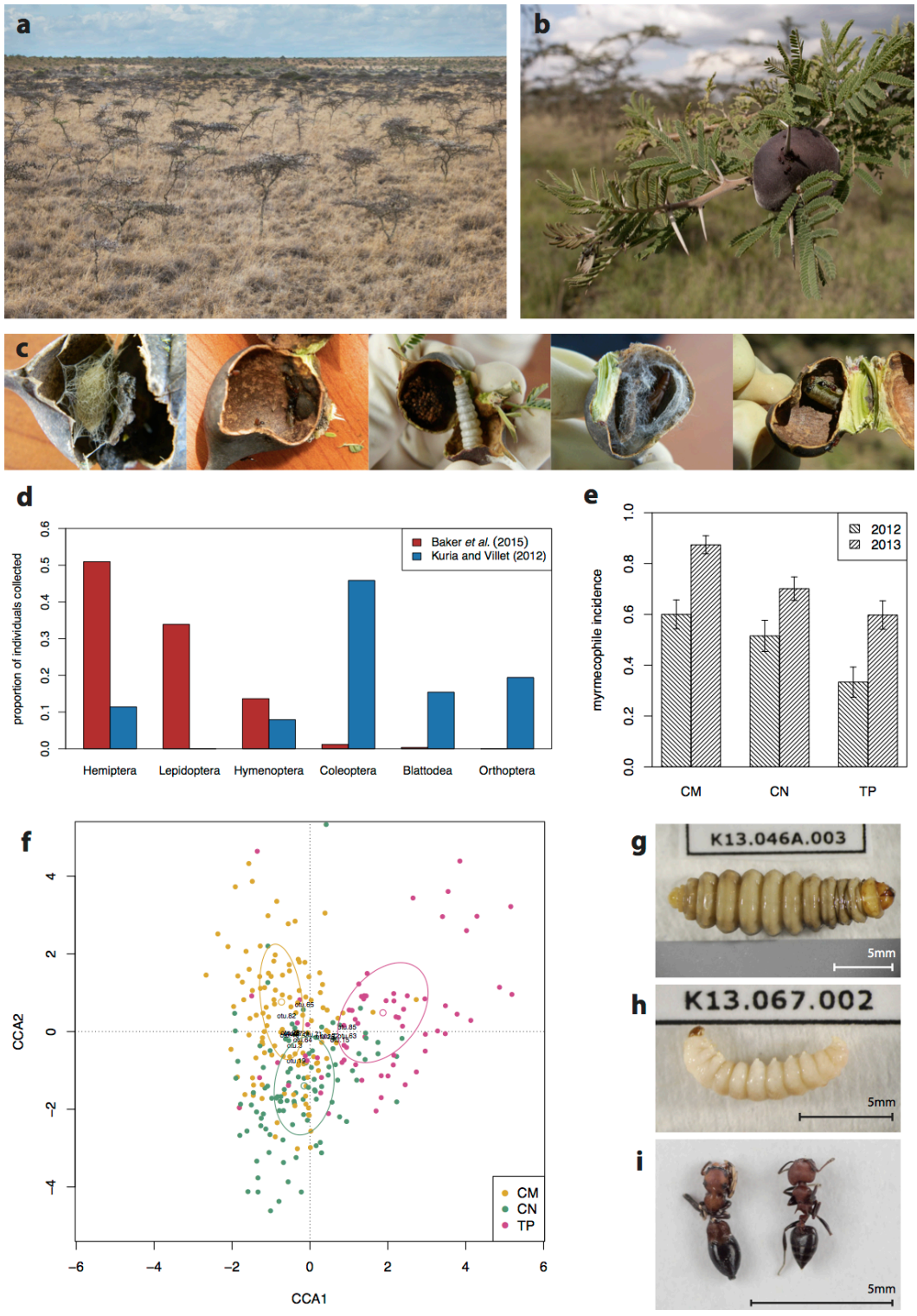

a
b
c
d
proportion of individuals collected
Baker et al. (2015)
Kuria and Villet (2012)
Hemiptera
Lepidoptera
Hymenoptera
Coleoptera
Blattodea
Orthoptera
e
myrmecophile incidence
2012
2013
CM
CN
TP
f
CCA2
CCA1
CM
CN
TP
g
K13.046A.003
5mm
h
K13.067.002
5mm
i
5mm

Figure A.1.

Figure A.1. Legend

(a) *V. drepanolobium* is typically the dominant tree species in East African black cotton savannas – virtually all trees visible in the image are *V. drepanolobium*.

(b) *V. drepanolobium* is covered with stipular thorns to defend against large mammalian herbivores. Many of the thorns are swollen and hollow, and serve as domatia inhabited by mutualistic ants.

(c) Many of the myrmecophiles in domatia of *V. drepanolobium* are immature forms that are difficult to identify using morphological characteristics. Molecular barcodes can be used to identify these myrmecophiles and link them to adult forms that are often better known or better described. Photo credit: Julianne Pelaez.

(d) Domatium myrmecophile communities (red bars, from [31]) are dominated by Hemiptera and Lepidoptera, but these taxa are less common in canopy insect communities (blue bars, from [25]). Domatium myrmecophiles also include spiders and snails, but these are omitted here for consistency with [25].

(e) Trees of *V. drepanolobium* occupied by colonies of *C. mimosae* tend to host more domatium myrmecophiles than trees occupied by colonies of *C. nigriceps*, which in turn host more than trees occupied by colonies of *T. penzigi*. From [31].

(f) Canonical correspondence analysis of myrmecophile communities showing that *C. mimosae*, *C. nigriceps* and *T. penzigi* ants associate with distinctive communities of domatium-dwelling myrmecophiles. Points on this plot, each representing a single tree, clearly separate according to the ant occupant, as denoted by colours of points. Black text shows the most abundant myrmecophile species in the same canonical correspondence space, with generalists in the centre of the plot and more specialized ant-associated taxa placed with the trees occupied by the ant species with which they commonly associate. Ellipsoids show standard deviations of the tree points.

(g) The tortricid moth, *Hystrichophora griseana,* is found on trees occupied by *C. mimosae* and *C. nigriceps* ants, but not on those occupied by *T. penzigi*.

(h) This gelechiid moth, *Dichomeris* sp., was found with all three ants.

(i) This salticid spider, *Myrmarachne* sp., (left) is a convincing visual mimic of *C. mimosae* ants (right), yet surprisingly was also found on trees occupied by all three ant species.

A wide range of myrmecophiles (‘ant lovers’) is also found living in the domatia alongside the ants. The ant-occupied domatia constitute a highly unique habitat – heavily defended by the ants against intruders, environmentally stable, and long-lived [30]. In response to this unique environment, we might expect domatium inhabitants in turn to be highly specialized. First, each of the domatium-dwelling ant species is highly aggressive, not only towards intruders that it detects, but also towards each other [16]. Myrmecophiles need to be able

to avoid the ants' defenses via mimicry, physical defenses, and/or engaging in mutualistic or manipulative interactions with the ants. We might therefore expect at least some myrmecophiles to specialize in their ant associations because of the degree of fine tuning required to interact successfully with their hosts. Second, we might expect some myrmecophiles to preferentially associate with one or more of the ant species if the ants differ in the benefits that they provide to the myrmecophiles, such as defense from predators. And third, we should see selection for lifestyles that capitalize on the stable and long-lived environment – for example, ant parasites with low costs and ant mutualists with low benefits [30] – and we thus expect domatium myrmecophile communities to be distinct from communities residing or transiently present in the canopies of the trees.

To explore the makeup of the domatium myrmecophile communities, we collected myrmecophiles exhaustively from 480 trees over two years at two sites in Kenya, for a total of 2361 individual myrmecophiles (see [31] for collection details). But deriving quantitative data from collections of domatium myrmecophiles is challenging. Many species are undescribed, and many of the myrmecophiles are immature forms that are often poorly known and difficult to identify (Figure A.1c). For example, out of the almost 600 individual Lepidoptera in our collection, 72.6% were larvae, 26.0% were pupae, and just 1.4% per cent were adults. DNA barcoding methods were therefore invaluable in examining these domatium myrmecophile communities, by serving in place of detailed morphological identifications [32].

We therefore sequenced COI barcodes for 1091 of our 2361 specimens in order to identify them. Since species-level taxonomic identifications were not always possible, we defined 'operational taxonomic units' (OTUs) for these specimens using the uclust clustering algorithm [33]. We classified a further 28 specimens based on visual inspection where we failed

to obtain good sequence. We also classified 1270 specimens that we did not sequence. These specimens belonged to 6 morphotypes, found with high abundance on a relatively small number of trees, for which standard COI barcode primers did not amplify (873 scale insects; see [34] but also see e.g. [35] for alternative primers) or for which the cost of sequencing all specimens did not appear to be justified (149 snails, 53 thrips, and 132 ants belonging to three taxa). The OTU-based classification of most specimens was not sensitive to the type of clustering algorithm or choice of similarity threshold. Nonetheless, for a small number of specimens, clustering choices did affect whether those specimens were grouped with others or classified as separate taxa, and we regard those specimens as good candidates for future investigation using molecular or morphological methods.

Our myrmecophile collections revealed that domatium communities were indeed taxonomically distinct from canopy communities (Figure A.1d and [31]). Domatium communities were dominated by Hemiptera and Lepidoptera, but these were less common in canopy insect communities, which were dominated by Coleoptera. (Domatium myrmecophiles also included many spiders and snails, but these were not reported for canopy communities in [25]).

As with the canopy insect communities [25], the abundance of domatium myrmecophiles differed among the ant species. Among the three domatium-dwelling ants, *C. mimosae* tended to host more myrmecophiles than *C. nigriceps*, which in turn hosted more than *T. penzigi* (Figure A.1e and [31]; see also [36]). Since most of the domatium myrmecophiles – particularly the Lepidoptera – are herbivorous [31], this pattern stands in contrast to the ants' defense against mammalian herbivores: *C. mimosae* is usually considered the best defender against large mammals, and *T. penzigi* the least effective.

Domatium myrmecophile communities also differed in composition among the ants. *C. mimosae*, *C. nigriceps* and *T. penzigi* ants tended to associate with distinctive communities of domatium dwelling myrmecophiles (Figure A.1f), though communities varied widely within each ant species. Some myrmecophiles showed strong specialization, as expected. Scale insects, for example, were almost always associated with *C. mimosae* ants. The tortricid moth *Hystrichophora griseana (*Figure A.1g) was very common with *C. mimosae* and *C. nigriceps*, but almost never found with *T. penzigi*. But for the most part, we found limited evidence for strong specialization on ant species. In many cases, this was because the number of individuals from an OTU was too small to clearly establish ant specialization. But there were also many cases where abundant taxa appeared to show no particular ant association. For example, notwithstanding strong specialization of *H. griseana*, many Lepidoptera (e.g. *Dichomeris* sp. in Figure A.1h) were associated with all three ants, and the use of DNA barcodes helps rule out the possibility that these are really cryptic species.

Perhaps the most surprising finding of generalist ant association was in the case of the abundant *Myrmarachne* sp. salticid spiders (Figure A.1i). Despite extremely strong visual mimicry of *C. mimosae* (*C. nigriceps* and *T. penzigi* ants differ in coloration, and *T. penzigi* also differs in body shape), these spiders were not found any more commonly with *C. mimosae* than with the other domatium dwelling ants. The visual mimicry probably plays no role in disguising spiders from the tree's ant residents, since most ants rely primarily on pheromones rather than visual cues to detect intruders. Instead, it probably serves to avoid predation by birds or parasitism by wasps. The close mimicry of *C. mimosae* combined with the fact that the spiders were found with ants other than *C. mimosae* suggests that selection favors mimicry of *C. mimosae* over other species of resident ant, presumably because they are the most bellicose

species, and that predators are deceived by the spiders' appearance as a worker of *C. mimosae*, but do not attend to the mismatch between the spiders and the tree's resident ants.

Our ongoing study of domatium myrmecophile communities has benefited greatly from DNA barcoding. The use of barcodes allows myrmecophiles to be collected and preserved rapidly in the field; identification of specimens across a broad taxonomic range can then easily be performed later by non-specialists. Although species-level taxonomic identifications are not always possible, especially in taxa that are not yet well represented in sequence databases, community-level patterns can still be analyzed by making use of similarity-based clustering into OTUs. Flexible, efficient and cost effective molecular protocols allow good throughput and thus increase feasibility for medium- to large-sized barcoding projects, in turn facilitating the detection of community-level patterns (e.g. automation-friendly DNA extractions [37]; we have also had good results with phenol-chloroform extractions on an AutoGenprep 965 robot, and with basic Chelex bead extractions [38]). While data management can be challenging for larger projects, we have found well-designed sequence processing pipelines (e.g. the Barcode of Life Data System [39]) and open-source relational database applications (e.g. VoSeq [40]) to be useful for managing sequences and other associated data.

Inquiline communities in *Nepenthes* pitchers

The aquatic pools enclosed by leaves of carnivorous pitcher plants contain communities of arthropods and microbes, and have been used for decades to study community dynamics [41-43]. Like the poorly known inhabitants of ant domatia, the protists and small organisms living in these pitchers can be difficult to identify by morphological methods alone. Metabarcoding, also known as next-generation amplicon sequencing [8, 44], is distinguished from conventional

barcoding by operating on the collective DNA of the environment rather than the isolated DNA of individual organisms. Thus, metabarcoding can provide a broader and less biased view of the organisms living within pitcher communities.

Pitchers of plants in the genus *Nepenthes* in Southeast Asia attract prey with extrafloral nectar, and have slippery edges and inner walls that cause insect visitors to fall in and drown [45]. The fluid inside the pitchers contains a mixture of rainwater and plant secretions. Pitchers catch and digest insect prey, but they also host thriving communities of aquatic arthropods, protozoa, bacteria, and fungi, called 'inquilines' [41]. Some inquilines have only been found living in *Nepenthes* pitchers, and appear to be endemic to these habitats [42]. To fully characterize and understand the communities within pitcher systems, we need a relatively unbiased view of the organisms present. Most previous studies of *Nepenthes* inquilines have been morphological [41-43], but in our recent study, we used metabarcoding to examine the eukaryotic communities within three *Nepenthes* (Figure A.2a) species in Singapore [46].

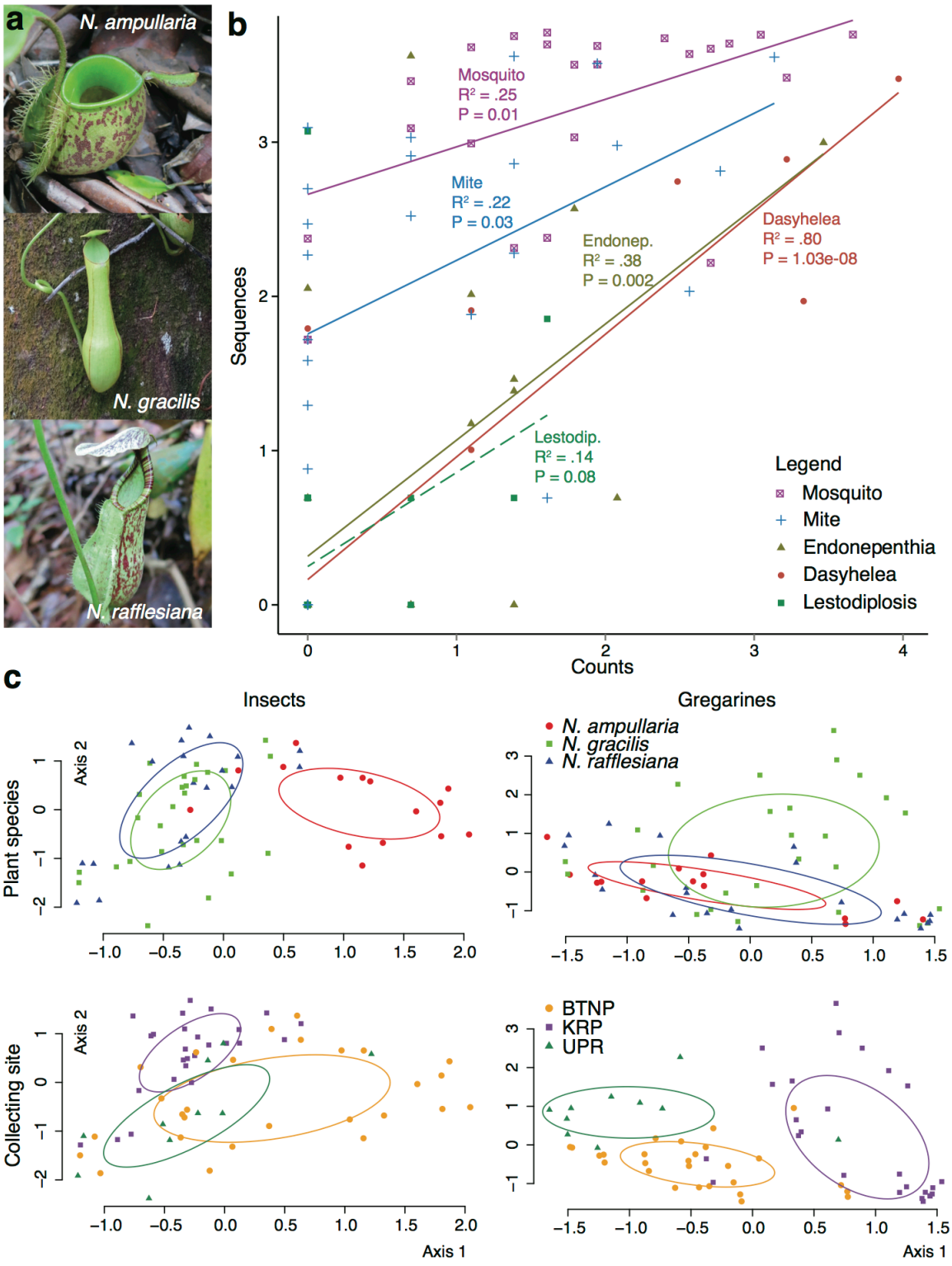

a
N. ampullaria
N. gracilis
N. rafflesiana
b
Mosquito
R² = .25
P = 0.01
Mite
R² = .22
P = 0.03
Endonep.
R² = .38
P = 0.002
Dasyhelea
R² = .80
P = 1.03e-08
Lestodip.
R² = .14
P = 0.08
Sequences
Counts
Legend
Mosquito
Mite
Endonepenthia
Dasyhelea
Lestodiplosis
c
Insects
Gregarines
Plant species
Collecting site
Axis 2
Axis 1
N. ampullaria
N. gracilis
N. rafflesiana
BTNP
KRP
UPR

Figure A.2.

Figure A.2. Legend

(a) Three species of *Nepenthes* pitcher plants studied in Bittleston *et al*. [46].

(b) Scatter plot of inquiline individual counts and 18S sequence counts plotted on a $\log_{10}$-$\log_{10}$ scale. Regression lines and *p*-values from the permutational linear models overlie the scatter plot. Solid lines are significant at $\alpha = 0.05$. Reproduced from Figure 3a of Bittleston *et al*. [46], copyright © 2015 by John Wiley Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.

(c) Non-metric multidimensional scale ordinations of insect communties (left panels) and gregarine communities (right panels). Each point is a different pitcher plant, coloured by pitcher plant species (top panels) and by collecting site (bottom panels). Ellipsoids are standard deviations of the points around the centres. Variation among insect communities is dominated by the difference between pitchers of *N. ampullaria* and pitchers of *N. gracilis* and *N. rafflesiana*. *N. ampullaria* is hypothesized to be more detritivorous than the other two species, which are predominantly carnivorous [50]. Variation among gregarine communities is dominated by variation among collecting sites.

An important question with metabarcoding is how relative abundances of sequences compare to actual organism counts, and whether community structure can be recovered [47]. Since *Nepenthes* pitchers are relatively self-contained, whole organism counts of insects can be compared to metabarcoding OTU counts. In our study, we therefore compared counts of inquiline insect larvae with metabarcoded 18S rDNA sequences from the same samples, to see how well the metabarcoding captured abundances of these organisms [46]. Positive correlations were found between the counts and sequences (Figure A.2b), suggesting that metabarcoding can be useful for understanding community structure of these organisms.

Our metabarcoding of *Nepenthes* pitchers also uncovered the presence of abundant gregarines (Apicomplexan protozoa), which are obligate parasites of invertebrates [46, 48]. The subclass Gregarinasina was the fourth most abundant eukaryotic group in *Nepenthes* pitchers, after insects, arachnids, and algae. Mosquito larvae have been shown to ingest gregarine oocytes while feeding [49]. The gregarines then complete their lifecycle in the mosquito midgut, and new oocysts are released into the environment during defecation, emergence into the adult form, or

upon ovoposition [49]. Morphological identification of gregarines in *Nepenthes* pitchers would be difficult, as they are small and can be hidden within the intestines of their hosts.

In our 18S metabarcoding study, insect OTUs largely mapped to dipteran inquilines living within the pitchers [46]. A bipartite network of insect and mite OTUs from the three *Nepenthes* plant species showed that these inquilines were significantly specialized. Certain OTUs tended to be found only within one plant species, while others were generalists found equally in all three hosts [46]. Insects adapted to the *Nepenthes* pitcher habitat might be able to distinguish between plant species, and adults might preferentially lay their eggs in certain species. Alternatively, conditions within the pitchers of different species may allow certain inquilines to thrive while inhibiting the growth of others.

Because gregarines are obligate parasites, and both insect and gregarine OTUs were in high abundances in pitcher habitats, we hypothesized that insect and gregarine diversity would follow similar patterns. To investigate this hypothesis, we performed a new analysis on the data from our previous publication [46] by separating the eukaryotic OTU table into insect and gregarine tables, and rarefying those tables to 1,922 and 200 sequences respectively. Twenty-one samples had fewer than 200 gregarine sequences and these samples were removed from both the insect and gregarine tables. We used distance-based redundancy analysis (dbRDA, function *capscale* in R) with Bray-Curtis distances to determine the effects of plant species, collection site, and collection year on insect and gregarine communities. On the dbRDA results, we used an ANOVA-like permutation test (function *anova* in the *vegan* package in R) with separate significance tests for each marginal term (plant species, collecting site, or collecting year) in a model with all other terms. Contrary to our expectations, insect and gregarine communities exhibited different drivers of diversity (Figure A.2c and Table A.1). For both taxa, the majority

of the variation was unexplained; however, a larger portion of the variation in insect community structure was explained by plant species ($p < 0.001$), while a larger portion of the variation in gregarine communities was explained by collecting site ($p < 0.001$). For both insects and gregarines, collection year did not significantly influence community structure.

Table A.1. Results of dbRDA analyses of *Nepenthes* inquilines

	Insects			Gregarines		
Term	**Variance**	**F**	**P**	**Variance**	**F**	**P**
Plant species	1.2055	3.3118	<0.001	0.8344	1.3719	0.108
Collection site	0.585	1.6071	0.069	2.8718	4.7217	<0.001
Collection year	0.1438	0.7899	0.587	0.4874	1.6026	0.086
Residual	10.3738			17.3341		

Table A.1. Results from distance-based redundancy analysis of *Nepenthes* insect communities (left) and gregarine communities (right) using Bray-Curtis distances, with host plant species, collection site, and collection year as predictors. Insect and gregarine communities had different correlates of diversity: plant species was a significant predictor for insect communities, while collection site was a significant predictor for gregarine communities.

Adult inquiline insects most likely can determine which plant host they are visiting when laying eggs, and we see that certain inquilines prefer certain plant species. The plant species-associated variation in insect communities seen in our new analysis primarily reflects a distinctive community in *N. ampullaria* relative to *N. gracilis* and *N. rafflesiana* (Figure A.2c). *N. ampullaria* is hypothesized to be less reliant on carnivory and more of a detritivore than other *Nepenthes* species [50]. The different ecology of *N. ampullaria* is potentially reflected in altered pitcher conditions, in turn selecting for different inquiline inhabitants.

Gregarine distributions, on the other hand, were better predicted by collection site than by host plant species. In our new analysis, the Kent Ridge Park (KRP) samples had gregarine communities that were more different than those from the other two sites, and geographically KRP is also farther away. Gregarine parasites could have been introduced into the pitchers via

adult inquilines, via prey species, or perhaps via abiotic vehicles such as raindrops. Considering the differences in diversity patterns of gregarines versus insects, we hypothesize that introduction via adult inquiline insects during oviposition is unlikely, as we would then have expected their distributions to be correlated. It is possible that gregarines could be encysted in a dormant stage within the pitcher fluid where they could utilize the assembly of many insects in one location to opportunistically infect new hosts. Alternatively, the gregarines may have complex infection and/or epidemiological dynamics with their host insects that we have yet to understand.

In general, insect inquilines appear to colonize *Nepenthes* pitchers more deterministically than gregarines, with certain organisms selecting specific host plant species, regardless of the location. Conversely, gregarines appear to colonize pitchers more stochastically, exhibiting a stronger correlation with collection site, an effect that could potentially be caused by some kind of dispersal limitation.

Metabarcoding provides a window into the complex interactions and patterns of biodiversity exhibited by pitcher plant systems. Barcode differences also help to discriminate between organisms (such as aquatic mites) that are often difficult to distinguish morphologically. Moreover, metabarcoding in this case has enabled us to identify microscopic gregarine parasites across multiple plant species and collection sites, and to uncover surprisingly different patterns of diversity between gregarines and insect inquilines. Barcodes are a valuable tool for generating and testing new hypothesis of community assembly, and can extend our investigations to organisms that are small and otherwise difficult to study.

Coevolutionary histories of animals and gut bacteria

The metabarcoding of host-associated microbial communities also has the potential to teach us something about the coevolutionary history of species relationships—interactions understood to be of major importance to a growing number of aspects of animal biology [51]. As with conventional barcoding of macrofauna, 16S rRNA gene-based barcoding of bacterial communities originated with the intent of identifying *which* taxa are present in a given environment. Since microbial taxonomy is still very incomplete [52], this typically involves the similarity-based clustering of 16S barcodes into 'operational taxonomic units,' or OTUs, to uncover patterns revealed by the distribution of these 'taxa' across hosts.

Host animals, unlike abiotic environments, themselves have an evolutionary history. It is widely appreciated that the distribution of microbial OTUs among hosts is a reflection of (and, possibly, an influence on) that evolutionary history [53, 54]: closely related animals frequently also host more similar microbial communities than do distant relatives [53, 55-57]. But these patterns of correlation between host phylogeny and microbial community similarity could result from a range of processes. Microbes could be inherited across host generations, resulting in codiversification of microbial lineages as a consequence of diversification in their hosts. Alternatively, related hosts could simply provide similar habitats, filtering similar microbes from the environment. These different processes also imply differences in the strength and nature of the effects host and microbe can have on each others' evolution. In part due to this ambiguous mapping from community pattern to evolutionary process, the question of how to interpret phylogenetic correlation in animal microbiota remains controversial [58].

Some additional insight into the origins of these correlations can be gleaned from consideration of metabarcode sequences not simply as taxonomic markers, but explicitly in light of their own evolutionary relationships. Metabarcode sequences reflect a phylogenetic history

that must be consistent with any proposed hypothesis for the origin of phylogenetic correlation, allowing us to place constraints on some of those hypotheses. For example, microbial diversification produced as a consequence of host diversification is constrained by the age of the host: consequently, the evolutionary distance between microbial barcodes in different hosts should have a recent upper bound if correlation between community similarity and host phylogeny arose via codiversification.

We can observe such a pattern in the gut microbial communities of South American turtle ants, in the genus *Cephalotes* (Figure A.3a). The diverse species of ants in this genus build their nests in empty cavities in trees and bushes, and host a dense gut microbiome that is thought to complement nutrient deficiencies in a largely herbivorous diet [59].

As has been reported in other systems, the gut microbiota of *Cephalotes* are correlated to host phylogeny (Figure A.3b). Using 454 metabarcoding of the bacterial 16S rRNA gene in guts from 25 *Cephalotes* species, we showed in a recent study that closely related ants also host more similar microbial communities [60]. But in the case of these ants, we were able to use the temporally structured evolutionary information within the barcodes themselves to give us some insight into how that similarity was likely to have arisen. Narrowing the similarity threshold used to define OTUs from the more typical 97% identity to 99% reveals the influence of more recent evolutionary history, splitting recently-diverged microbial lineages that would have been collapsed into single OTUs at the wider threshold. Doing so increases the separation apparent between clades of related hosts in a network visualization of these communities (Figure A.3c). Wider OTU definitions also obscure correlations between clustering dendrograms of community similarity metrics and host phylogeny (Figure 3 A.d) that are apparent at narrower definitions (Figure A.3e). That such phylogenetic correlation is only apparent when considering information

about relatively recent bacterial evolution is consistent with it being generated through processes like codiversification or phylogenetically restricted host shifts [61].

If codiversification does explain the similarity of communities from related host species, we should also be able to see a signal of host phylogeny in metabarcode sequences from individual microbial lineages. At least to some extent, we do (Figure A.3f). Taking advantage of the structure of diversity in the *Cephalotes* gut, we performed an additional analysis of our metabarcode data from [60] to examine a lineage of Verrucomicrobia that is both universally present and abundant in these communities, and for which there is usually only one dominant strain per host community. We took the representative 16S metabarcode sequence for the 99% OTU assigned to the Verrucomicrobia lineage that was most abundant in each *Cephalotes* colony, aligned all extracted sequences using MUSCLE and then built a pseudo-maximum likelihood phylogeny of these barcodes using FastTree. A tanglegram analysis of this bacterial tree shows substantial but imperfect correlation with host phylogeny, suggesting that this lineage may indeed be codiversifying with the host. That this correlation is weaker than the aggregate signal for the entire community (Figure A.3e) further suggests that other lineages in the community are undergoing similar processes.

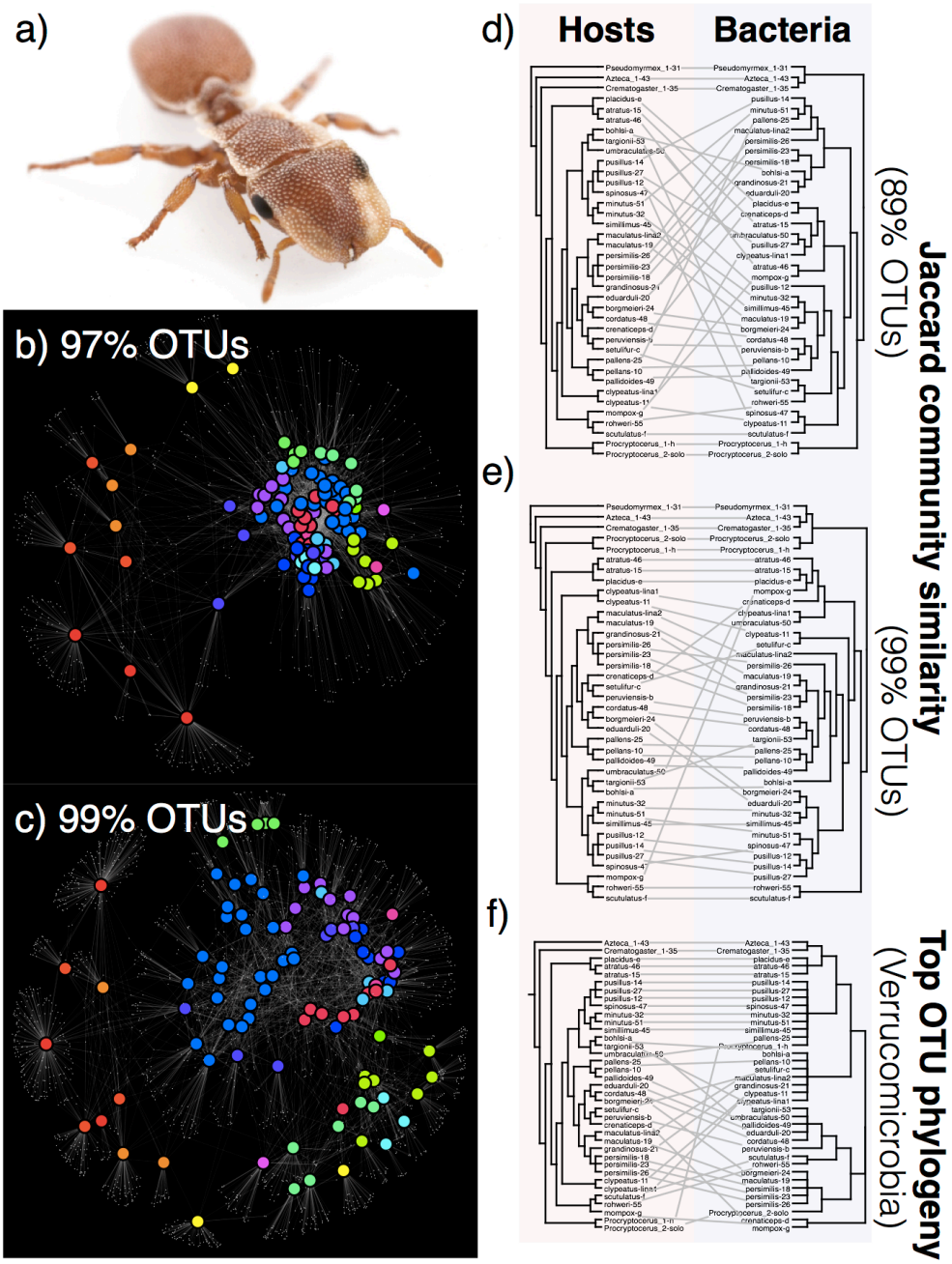

a)
b) 97% OTUs
c) 99% OTUs
d)
Hosts
Bacteria
e)
f)
Jaccard community similarity
(89% OTUs)
(99% OTUs)
Top OTU phylogeny
(Verrucomicrobia)

Figure A.3.

Figure A.3. Legend:

(a) *Cephalotes* ant.

(b) 97% identity Operational Taxonomic Unit (OTU) network visualization of *Cephalotes* gut microbiota. Host samples (colored icons) are connected by edges to 97% OTUs (small white dots). *Cephalotes* samples are colored by host clade, with closely related species having similar colors. *Cephalotes* microbiota group together separately from non-*Cephalotes* microbiota (yellow, orange, and red colors), as they share a large number of OTUs.

(c) Identical to (b) except that OTUs are defined at 99% identity. Note that at 99% identity, far fewer OTUs are shared by all *Cephalotes* microbiota, but samples from related species continue to group together.

(d) Tanglegram linking cladograms of host phylogenetic relationships with microbiota similarity relationships, defined by UPGMA clustering of Jaccard dissimilarities among *Cephalotes* colonies calculated with 89% OTUs. At this level, separation of *Cephalotes* microbiota from non-*Cephalotes* microbiota is retained, but phylogenetic relationships within *Cephalotes* are not reflected in microbiota similarities.

(e) As in (d), except using 99% OTUs. At this level, similarity among microbiota also reflects phylogenetic relationships within *Cephalotes*.

(f) Tanglegram linking phylogeny of the highest-abundance member of the Verrucomicrobia from each *Cephalotes* colony to phylogeny of the hosts.

In principle, such lineage-by-lineage analyses offer the potential to sift through whole communities to identify the specific microbes shaping phylogenetic correlation in microbiomes—giving us a potentially powerful tool for understanding these complex systems. Separating lineages by their evolutionary fidelity to hosts could help to identify microbes especially likely to be of functional import, whether due to explicit reciprocal coevolution with the host or simply as a byproduct of having been a constant element of the host's internal environment.

In practice, limitations in typical metabarcoding approaches prevent drawing such conclusions with high sensitivity or specificity. The 16S rRNA gene evolves slowly. With the relatively short read lengths of current Illumina and Ion Torrent platforms, even tens of millions of years of divergence may only be supported by a handful of phylogenetically informative

characters, resulting in poor phylogenetic reconstructions. Sequencing error further obscures this pattern.

Still, interrogation of the evolutionary history represented in metabarcode sequences has yielded a number of interesting cases, especially when combined with other techniques to increase the amount of useful information available for analysis. In bumble bees [62] and pyrrhocorid seed bugs [63], low-throughput follow-up sequencing of target lineages using specific primers permitted deeper exploration of trends observed in untargeted metabarcoding efforts. In vertebrates, techniques to reduce the impact of sequencing noise permitted the detection of patterns of host specificity from metabarcoding data, even though the underlying sequences were quite similar [55].

As new sequencing approaches are developed, analysis of evolutionary history directly from metabarcode data will become possible with more confidence. Long-read technology will allow the use of full-length gene sequences, provided current problems of read accuracy can be overcome. Even given current sequencing technology, changing the bacterial metabarcoding target to faster-evolving protein-coding genes will yield more phylogenetically informative information than the 16S gene. Recent work has already made this approach possible, either by the initial amplification of these genes [64] or by what is effectively post-hoc barcoding of microbial communities by sifting through shotgun metagenomic sequence data [65, 66].

Conclusion

Our studies illustrate the value of DNA barcoding and metabarcoding for identifying taxa in host-symbiont community interactions. For organisms like myrmecophiles (often juvenile invertebrates) on *V. drepanolobium*, barcoding has provided us with a way to identify specimens

that would otherwise be difficult to classify. Metabarcoding methods likewise have allowed us to detect and identify inquiline taxa in *Nepenthes* pitcher plants, and gut bacterial symbionts in *Cephalotes* ants.

But our studies also show how the utility of DNA barcodes can extend beyond the simple identification of individual symbionts, to the examination of ecological patterns [67]. This in part reflects the relatively high sample throughput permitted by barcoding methods, which facilitates the accurate profiling of entire communities, and offers the opportunity to assess interactions among symbionts and to identify patterns such as ecological convergence that may emerge only at the community level [68]. In our *V. drepanolobium* and *Nepenthes* studies, this high throughput was primarily realized through efficiencies in sample collection, sample processing, and data analysis. As technology improves, an additional efficiency will become increasingly relevant: the availability of rapid in-the-field sequencing, using portable devices such as the Oxford Nanopore MinION [69], will permit almost real-time feedback on specimens and environmental samples. This will allow researchers to refine sample- and data collection strategies on the fly (e.g. what are appropriate sample sizes and spatial scales for sampling?), and to generate new hypotheses that can be tested immediately instead of having to wait until the next field trip.

Although the value of DNA barcodes for species discovery and delineation has been challenged (e.g. [70]), we have found OTU clustering of arthropods and protozoa to be reasonably robust to choices of algorithm or parameters in both our *V. drepanolobium* and *Nepenthes* studies. Where results are sensitive to clustering choices, however, we are happy to adopt a relaxed approach to barcoding, and flag those specimens for further investigation using other markers or morphology, rather than rely solely on our barcoding data. Our analyses of

differences in symbiont community composition between different host species largely sidestep uncertainty in the taxonomic placement or phylogenetic relationships of our OTUs: our analyses demonstrate community differences based on OTU abundances for each host, but not on the taxonomic labels attached to those OTUs, or on their phylogenetic placements (cf. [71]).

DNA barcodes can also provide a window on the evolutionary history of a host-symbiont association – a dynamic relationship shaped by selection and phylogenetic constraint that is absent in abiotic contexts. This reflects the fact that barcodes are not just taxonomic labels, but evolving DNA sequences that can be analyzed for evidence of host-symbiont codiversification. In contrast to our *V. drepanolobium* and *Nepenthes* studies, the clustering of *Cephalotes* gut bacteria is sensitive to our choice of clustering threshold. But rather than being problematic, we are in fact able to use hierarchical clustering at different thresholds to our advantage, interpreting this sensitivity to parameters in light of expectations about the timescales of coevolutionary change.

As technology and methods improve, barcoding and metabarcoding approaches will become increasingly useful for ecological and evolutionary studies. Longer sequence reads and lower error rates, for example, will increase our capacity to draw inferences especially regarding recent phylogenetic history. The development of a wider range of sequencing targets will also help make barcoding approaches useful for a wider range of organisms and research questions. Indeed, as sequencing becomes cheaper, metagenomic datasets will allow appropriate barcode markers to be chosen *ex post* [65, 66], or even for diversity assessments based on genome assemblies [72]. These approaches need not replace the simplicity of a single, standardized barcode region [32], but should nonetheless be embraced as a valuable expansion of the barcoding approach [73].

DNA barcodes will also become increasingly useful for ecological and evolutionary studies as sequence and other data accumulate in public databases. As these databases expand, we need to ensure that the widest possible selection of data can be accessed in an automated fashion, by encouraging researchers to annotate published data with as much machine-readable metadata as possible. Location, habitat or timestamp data on DNA barcodes, for example, may help generate more accurate pictures of species distributions over space and time, and the ecological correlates of those distributions. Conservation and barcode data can be combined in order to generate phylogenetically-informed conservation assessments [74]. But barcodes and barcode-based taxon assignments also represent a natural and convenient way to connect a wide range of data from different datasets: images and information from museum and library digitization projects, location and other metadata from collections, morphological information, natural history observations, stable isotope data, or even data on metabolic rates [75]. Combining datasets potentially allows researchers to uncover patterns across larger temporal, spatial or phylogenetic scales than would normally be feasible [67]. Combining multiple data types – e.g. on symbiont community composition, on genomic functional capacities, and on the nature of trophic or other interactions among organisms – potentially allows us to, for example, identify emergent properties of communities or rules governing the assembly of symbiont communities [68], or to assess changes in community structure that might act as signals of ecological distress [76]. Connecting many disparate datasets so they are inter-referential is not a trivial challenge, but one that holds great potential for furthering our understanding of species interactions.

Appendix A Acknowledgments

We thank Paul Hebert for his inspiring and indefatigable efforts in spearheading the International Barcode of Life (*iBOL*), as well as Mehrdad Hajibabaei and Peter Hollingsworth for inviting us to participate in this special issue. We are grateful to the three anonymous reviewers whose comments have helped us improve the manuscript.

Ethics

Not applicable.

Authors' Contributions

CCMB and NEP conceived of the general theme of the paper; CCMB collected data and interpreted results relating to *V. drepanolobium*; LSB collected data and interpreted results relating to *Nepenthes*; and JGS collected data and interpreted results relating to *Cepahalotes*. CCMB and NEP organized and drafted the initial outline, and all authors wrote and revised the manuscript.

Competing Interests

We have no competing interests.

Funding

CCMB, LSB and JGS were supported by fellowships from the Department of Organismic and Evolutionary Biology at Harvard University, and by Putnam Expedition Grants from Harvard's Museum of Comparative Zoology. LSB and JGS were supported by NSF Graduate

Research Predoctoral Fellowships and NSF Doctoral Dissertation Improvement Grants DEB-1400982 and DEB-11105515 respectively. CCMB and NEP were supported by a Foundational Questions in Evolutionary Biology (FQEB) grant from the Templeton Foundation and by NSF SES-0750480 as part of the European Science Foundation's call for research on The Evolution of Cooperation and Trading (TECT).

Appendix A References

[1] Breznak, J.A. & Brune, A. 1994 Role of microorganisms in the digestion of lignocellulose by termites. *Annual Review of Entomology* **39**, 453-487. (doi:10.1146/annurev.en.39.010194.002321).

[2] Rettenmeyer, C.W., Rettenmeyer, M.E., Joseph, J. & Berghoff, S.M. 2011 The largest animal association centered on one species: the army ant *Eciton burchellii* and its more than 300 associates. *Insectes Sociaux* **58**, 281-292. (doi:10.1007/s00040-010-0128-8).

[3] Hodkinson, B.P. & Lutzoni, F. 2009 A microbiotic survey of lichen-associated bacteria reveals a new lineage from the Rhizobiales. *Symbiosis* **49**, 163-180. (doi:10.1007/s13199-009-0049-3).

[4] Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. 2004 Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *PNAS* **101**, 14812-14817. (doi:10.1073/pnas.0406166101).

[5] Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S. & Francis, C.M. 2004 Identification of Birds through DNA Barcodes. *PLoS Biol* **2**, e312. (doi:10.1371/journal.pbio.0020312).

[6] Dincă, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M.L., García-Barros, E., Hebert, P.D.N. & Vila, R. 2015 DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports* **5**, 12395. (doi:10.1038/srep12395).

[7] Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R. & Abebe, E. 2005 Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1935-1943. (doi:10.1098/rstb.2005.1725).

[8] Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. 2012 Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21**, 2045-2050. (doi:10.1111/j.1365-294X.2012.05470.x).

[9] Clare, E.L., Fraser, E.E., Braid, H.E., Fenton, M.B. & Hebert, P.D.N. 2009 Species on the menu of a generalist predator, the eastern red bat (Lasiurus borealis): using a molecular approach to detect arthropod prey. *Molecular Ecology* **18**, 2532-2542. (doi:10.1111/j.1365-294X.2009.04184.x).

[10] Kartzinel, T.R., Chen, P.A., Coverdale, T.C., Erickson, D.L., Kress, W.J., Kuzmina, M.L., Rubenstein, D.I., Wang, W. & Pringle, R.M. 2015 DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *PNAS* **112**, 8019-8024. (doi:10.1073/pnas.1503283112).

[11] Derocles, S.A.P., Evans, D.M., Nichols, P.C., Evans, S.A. & Lunt, D.H. 2015 Determining Plant – Leaf Miner – Parasitoid Interactions: A DNA Barcoding Approach. *PLoS ONE* **10**, e0117872. (doi:10.1371/journal.pone.0117872).

[12] Nyman, T., Leppänen, S.A., Várkonyi, G., Shaw, M.R., Koivisto, R., Barstad, T.E., Vikberg, V. & Roininen, H. 2015 Determinants of parasitoid communities of willow-galling sawflies: habitat overrides physiology, host plant and space. *Molecular Ecology* **24**, 5059-5074. (doi:10.1111/mec.13369).

[13] Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. 2007 Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature* **445**, 202-205. (doi:10.1038/nature05429).

[14] Young, T.P., Stubblefield, C.H. & Isbell, L.A. 1997 Ants on swollen-thorn acacias: species coexistence in a simple system. *Oecologia* **109**, 98-107. (doi:10.1007/s004420050063).

[15] Hocking, B. 1970 Insect associations with the swollen thorn acacias. *Transactions of the Royal Entomological Society of London* **122**, 211-255. (doi:10.1111/j.1365-2311.1970.tb00532.x).

[16] Palmer, T.M., Young, T.P., Stanton, M.L. & Wenk, E. 2000 Short-term dynamics of an acacia ant community in Laikipia, Kenya. *Oecologia* **123**, 425-435. (doi:10.1007/s004420051030).

[17] Janzen, D.H. 1985 The Natural History of Mutualisms. In *The Biology of Mutualism* (ed. D.H. Boucher), pp. 40-99. New York, Oxford University Press.

[18] Madden, D. & Young, T.P. 1992 Symbiotic ants as an alternative defense against giraffe herbivory in spinescent *Acacia drepanolobium*. *Oecologia* **91**, 235-238. (doi:10.1007/BF00317789).

[19] Stapley, L. 1998 The interaction of thorns and symbiotic ants as an effective defence mechanism of swollen-thorn acacias. *Oecologia* **115**, 401-405. (doi:10.1007/s004420050534).

[20] Goheen, J.R. & Palmer, T.M. 2010 Defensive Plant-Ants Stabilize Megaherbivore-Driven Landscape Change in an African Savanna. *Current Biology* **20**, 1768-1772. (doi:10.1016/j.cub.2010.08.015).

[21] Palmer, T.M. & Brody, A.K. 2007 Mutualism as reciprocal exploitation: African plant-ants defend foliar but not reproductive structures. *Ecology* **88**, 3004-3011. (doi:10.1890/07-0133.1).

[22] Martins, D.J. 2010 Not all ants are equal: obligate acacia ants provide different levels of protection against mega-herbivores. *African Journal of Ecology* **48**, 1115-1122. (doi:10.1111/j.1365-2028.2010.01226.x).

[23] Stanton, M.L., Palmer, T.M., Young, T.P., Evans, A. & Turner, M.L. 1999 Sterilization and canopy modification of a swollen thorn acacia tree by a plant-ant. *Nature* **401**, 578-581. (doi:10.1038/44119).

[24] Palmer, T., Young, T. & Stanton, M. 2002 Burning bridges: priority effects and the persistence of a competitively subordinate acacia-ant in Laikipia, Kenya. *Oecologia* **133**, 372-379. (doi:10.1007/s00442-002-1026-1).

[25] Kuria, S.K. & Villet, M.H. 2012 The role of ants and mammalian herbivores on the structure and composition of insect communities found on canopies of *Acacia drepanolobium*. *African Journal of Agricultural Research* **7**, 5317-5331. (doi:10.5897/AJAR12.059).

[26] Monod, T. & Schmitt, C. 1968 Contribution à l'étude des pseudo-galles formicaires chez quelques Acacias africains. *Bulletin de l'Institut Fondamental d'Afrique Noire - Série A, Sciences Naturelles* **30**, 953-1012.

[27] Martins, D.J., Collins, S.C., Congdon, C. & Pierce, N.E. 2013 Association between the African lycaenid, *Anthene usamba*, and an obligate acacia ant, *Crematogaster mimosae*. *Biological Journal of the Linnean Society* **109**, 302-312. (doi:10.1111/bij.12037).

[28] Quicke, D.L.J. & Stanton, M.L. 2005 *Trigastrotheca laikipiensis* sp. nov. (Hymenoptera: Braconidae): A New Species of Brood Parasitic Wasp that Attacks Foundress Queens of Three Coexisting Acacia-ant Species in Kenya. *Journal of Hymenoptera Research* **14**, 182-190.

[29] Schumer, M., Birger, R., Tantipathananandh, C., Aurisano, J., Maggioni, M. & Mwangi, P. 2013 Infestation by a Common Parasite is Correlated with Ant Symbiont Identity in a Plant-Ant Mutualism. *Biotropica* **45**, 276-279. (doi:10.1111/btp.12038).

[30] Hughes, D.P., Pierce, N.E. & Boomsma, J.J. 2008 Social insect symbionts: evolution in homeostatic fortresses. *Trends in Ecology and Evolution* **23**, 672-677. (doi:10.1016/j.tree.2008.07.011).

[31] Baker, C.C. 2015 Complexity in mutualisms: indirect interactions with multiple parties [PhD dissertation], Harvard University.

[32] Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. 2003 Biological identifications through DNA barcodes. *Proc R Soc B* **270**, 313-321. (doi:10.1098/rspb.2002.2218).

[33] Edgar, R.C. 2010 Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461. (doi:10.1093/bioinformatics/btq461).

[34] Kondo, T., Gullan, P.J. & Williams, D.J. 2008 Coccidology: The study of scale insects (Hemiptera: Sternorrhyncha: Coccoidea). *Revista Corpoica - Ciencia y Tecnologia Agropecuaria* **9**, 55-61.

[35] Park, D.-S., Suh, S.-J., Oh, H.-W. & Hebert, P.D.N. 2010 Recovery of the mitochondrial COI barcode region in diverse Hexapoda through tRNA-based primers. *BMC Genomics* **11**, 423-423. (doi:10.1186/1471-2164-11-423).

[36] Isbell, L.A. & Young, T.P. 2007 Interspecific and temporal variation of ant species within *Acacia drepanolobium* ant domatia, a staple food of patas monkeys (*Erythrocebus patas*) in Laikipia, Kenya. *Am J Primatol* **69**, 1387-1398. (doi:10.1002/ajp.20444).

[37] Ivanova, N.V., Dewaard, J.R. & Hebert, P.D.N. 2006 An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes* **6**, 998-1002. (doi:10.1111/j.1471-8286.2006.01428.x).

[38] Gadau, J.r. 2009 DNA Isolation from Ants. *Cold Spring Harb Protoc* **4**. (doi:10.1101/pdb.prot5245).

[39] Ratnasingham, S. & Hebert, P.D.N. 2007 BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes* **7**, 355-364. (doi:10.1111/j.1471-8286.2007.01678.x).

[40] Peña, C. & Malm, T. 2012 VoSeq: A Voucher and DNA Sequence Web Application. *PLoS ONE* **7**, e39071. (doi:10.1371/journal.pone.0039071).

[41] Beaver, R.A. 1983 The communities living in *Nepenthes* pitcher plants: fauna and food webs. In *Phytotelmata: Plants as hosts for aquatic insect communities* (eds. J. Frank & L. Lounibos), pp. 129-159. New Jersey, Plexus Publishing.

[42] Kitching, R.L. 2000 *Food webs and container habitats: The Natural History and Ecology of Phytotelmata*. New York, Cambridge University Press.

[43] Clarke, C.M. & Kitching, R.L. 1993 The metazoan food webs from six Bornean *Nepenthes* species. *Ecological Entomology* **18**, 7-16. (doi:10.1111/j.1365-2311.1993.tb01074.x).

[44] Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W. & Huse, S.M. 2009 A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE* **4**, e6372. (doi:10.1371/journal.pone.0006372).

[45] Juniper, B.E., Robins, R.J. & Joel, D.M. 1989 *The Carnivorous Plants*. London ; San Diego, Academic Press.

[46] Bittleston, L.S., Baker, C.C.M., Strominger, L.B., Pringle, A. & Pierce, N.E. 2015 Metabarcoding as a tool for investigating arthropod diversity in *Nepenthes* pitcher plants. *Austral Ecology* **Accepted for publication April 2015**. (doi:10.1111/aec.12271).

[47] Amend, A.S., Seifert, K.A. & Bruns, T.D. 2010 Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Molecular Ecology* **19**, 5555-5565. (doi:10.1111/j.1365-294X.2010.04898.x).

[48] Desportes, I. & Schrével, J. (eds.). 2013 The Gregarines: The Early Branching Apicomplexa. Leiden, Brill.

[49] Roychoudhury, S. & Kobayashi, M. 2006 New findings on the developmental process of *Ascogregarina taiwanensis* and *Ascogregarina culicis* in *Aedes albopictus* and *Aedes aegypti*. *Journal of the American Mosquito Control Association* **22**, 29-36. (doi:10.2987/8756-971X(2006)22[29:NFOTDP]2.0.CO;2).

[50] Moran, J.A., Clarke, C.M. & Hawkins, B.J. 2003 From carnivore to detritivore? Isotopic evidence for leaf litter utilization by the tropical pitcher plant *Nepenthes ampullaria*. *Int J Plant Sci* **164**, 635-639. (doi:10.1086/375422).

[51] McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S.F., et al. 2013 Animals in a bacterial world, a new imperative for the life sciences. *PNAS* **110**, 3229-3236. (doi:10.1073/pnas.1218525110).

[52] Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., et al. 2014 Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431-437. (doi:10.1038/nature12352).

[53] Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., et al. 2008 Evolution of Mammals and Their Gut Microbes. *Science* **320**, 1647-1651. (doi:10.1126/science.1155725).

[54] van Opstal, E.J. & Bordenstein, S.R. 2015 Rethinking heritability of the microbiome. *Science* **349**, 1172-1173. (doi:10.1126/science.aab3958).

[55] Reveillaud, J., Maignien, L., Eren, A.M., Huber, J.A., Apprill, A., Sogin, M.L. & Vanreusel, A. 2014 Host-specificity among abundant and rare taxa in the sponge microbiome. *ISME J* **8**, 1198-1209. (doi:10.1038/ismej.2013.227).

[56] Abdul Rahman, N., Parks, D., Willner, D., Engelbrektson, A., Goffredi, S., Warnecke, F., Scheffrahn, R. & Hugenholtz, P. 2015 A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. *Microbiome* **3**, 1-16. (doi:10.1186/s40168-015-0067-8).

[57] Ochman, H., Worobey, M., Kuo, C.-H., Ndjango, J.-B.N., Peeters, M., Hahn, B.H. & Hugenholtz, P. 2010 Evolutionary Relationships of Wild Hominids Recapitulated by Gut Microbial Communities. *PLoS Biol* **8**, e1000546. (doi:10.1371/journal.pbio.1000546).

[58] Moran, N.A. & Sloan, D.B. 2015 The Hologenome Concept: Helpful or Hollow? *PLoS Biology* **13**, e1002311. (doi:10.1371/journal.pbio.1002311).

[59] Hu, Y., Łukasik, P., Moreau, C.S. & Russell, J.A. 2014 Correlates of gut community composition across an ant species (*Cephalotes varians*) elucidate causes and consequences of symbiotic variability. *Molecular Ecology* **23**, 1284-1300. (doi:10.1111/mec.12607).

[60] Sanders, J.G., Powell, S., Kronauer, D.J.C., Vasconcelos, H.L., Frederickson, M.E. & 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).

[61] De Vienne, D.M., Giraud, T. & Shykoff, J.A. 2007 When can host shifts produce congruent host and parasite phylogenies? A simulation approach. *Journal of Evolutionary Biology* **20**, 1428-1438. (doi:10.1111/j.1420-9101.2007.01340.x).

[62] Koch, H., Abrol, D.P., Li, J. & Schmid-Hempel, P. 2013 Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. *Molecular Ecology* **22**, 2028-2044. (doi:10.1111/mec.12209).

[63] Sudakaran, S., Retz, F., Kikuchi, Y., Kost, C. & Kaltenpoth, M. 2015 Evolutionary transition in symbiotic syndromes enabled diversification of phytophagous insects on an imbalanced diet. *ISME J* **9**, 2587-2604. (doi:10.1038/ismej.2015.75).

[64] Caro-Quintero, A. & Ochman, H. 2015 Assessing the Unseen Bacterial Diversity in Microbial Communities. *Genome Biology and Evolution*, evv234. (doi:10.1093/gbe/evv234).

[65] Darling, A.E., Jospin, G., Lowe, E., Matsen, F.A., Bik, H.M. & Eisen, J.A. 2014 Phylosift: Phylogenetic analysis of genomes and metagenomes. *PeerJ* **2**, e243. (doi:10.7717/peerj.243).

[66] Nguyen, N.-p., Mirarab, S., Liu, B., Pop, M. & Warnow, T. 2014 TIPP: Taxonomic Identification and Phylogenetic Profiling. *Bioinformatics* **2014**. (doi:10.1093/bioinformatics/btu721).

[67] Joly, S., Davies, T.J., Archambault, A., Bruneau, A., Derry, A., Kembel, S.W., Peres-Neto, P., Vamosi, J. & Wheeler, T.A. 2014 Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Mol Ecol Resour* **14**, 221-232. (doi:10.1111/1755-0998.12173).

[68] Bittleston, L.S., Pierce, N.E., Ellison, A.M. & Pringle, A. Convergence in Multispecies Interactions. *Trends in Ecology & Evolution*. (doi:10.1016/j.tree.2016.01.006).

[69] Mikheyev, A.S. & Tin, M.M.Y. 2014 A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour* **14**, 1097-1102. (doi:10.1111/1755-0998.12324).

[70] Rubinoff, D., Cameron, S. & Will, K. 2006 A Genomic Perspective on the Shortcomings of Mitochondrial DNA for "Barcoding" Identification. *Journal of Heredity* **97**, 581-594. (doi:10.1093/jhered/esl036).

[71] Lozupone, C. & Knight, R. 2005 UniFrac: A New Phylogenetic Method for Comparing Microbial Communities. *Applied Environmental Microbiology* **71**, 8228-8235. (doi:10.1128/AEM.71.12.8228-8235.2005).

[72] Tang, M., Hardman, C.J., Ji, Y., Meng, G., Liu, S., Tan, M., Yang, S., Moss, E.D., Wang, J., Yang, C., et al. 2015 High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution* **6**, 1034-1043. (doi:10.1111/2041-210X.12416).

[73] Taylor, H.R. & Harris, W.E. 2012 An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Mol Ecol Resour* **12**, 377-388. (doi:10.1111/j.1755-0998.2012.03119.x).

[74] Guralnick, R. & Hill, A. 2009 Biodiversity informatics: automated approaches for documenting global biodiversity patterns and processes. *Bioinformatics* **25**, 421-428. (doi:10.1093/bioinformatics/btn659).

[75] April, J., Hanner, R.H., Mayden, R.L. & Bernatchez, L. 2013 Metabolic Rate and Climatic Fluctuations Shape Continental Wide Pattern of Genetic Divergence and Biodiversity in Fishes. *PLoS ONE* **8**, e70296. (doi:10.1371/journal.pone.0070296).

[76] Tylianakis, J.M., Laliberté, E., Nielsen, A. & Bascompte, J. 2010 Conservation of species interaction networks. *Biological Conservation* **143**, 2270-2279. (doi:10.1016/j.biocon.2009.12.004).

Supplementary Material from Chapter 2:

Supplementary Table 2.1. *Nepenthes* samples and metadata

Sample	Year	Host Species	Coll. Site	pH	Tot. Vol. (mL)	Pooled	Pre-Extraction	Cleaned
N01B	2012	gracilis	BNTP	NA	NA	No	Filter	No
N02A	2012	ampullaria	BNTP	NA	NA	No	Filter	No
N02B	2012	ampullaria	BNTP	NA	NA	No	Filter	No
N03C	2012	ampullaria	BNTP	NA	NA	No	Filter	No
N06C	2012	gracilis	BNTP	NA	NA	No	Filter	No
N07A	2012	ampullaria	BNTP	NA	NA	No	Filter	No
N07B	2012	ampullaria	BNTP	NA	NA	No	Filter	No
N08A	2012	rafflesiana	BNTP	NA	NA	No	Filter	No
N10B.	2012	rafflesiana	BNTP	NA	NA	No	Filter	No
N12	2012	rafflesiana	UPR	NA	NA	No	Filter	No
N13	2012	gracilis	UPR	NA	NA	Yes	Filter	No
N15	2012	rafflesiana	UPR	NA	NA	No	Filter	No
N20	2012	gracilis	KRP	NA	NA	Yes	Filter	No
N21	2012	gracilis	KRP	NA	NA	Yes	Filter	No
N22	2012	gracilis	KRP	NA	NA	Yes	Filter	No
N23A	2012	gracilis	KRP	NA	NA	No	Filter	No
N23B	2012	gracilis	KRP	NA	NA	No	Filter	No
N24	2012	gracilis	KRP	NA	NA	No	Filter	No
N27A	2012	rafflesiana	KRP	NA	NA	No	Filter	No
N27C	2012	rafflesiana	KRP	NA	NA	No	Filter	No
N28	2012	rafflesiana	KRP	NA	NA	No	Filter	No
N29A	2012	rafflesiana	KRP	NA	NA	No	Filter	No
N32	2012	gracilis	KRP	NA	NA	Yes	Filter	No
N33	2012	rafflesiana	KRP	NA	NA	No	Filter	No
N34	2012	rafflesiana	KRP	NA	NA	No	Filter	No
N37A.	2012	ampullaria	KRP	NA	NA	No	Filter	No
N37B	2012	ampullaria	KRP	NA	NA	No	Filter	No
N38B.	2012	ampullaria	KRP	NA	NA	No	Filter	No
Ne02	2013	gracilis	KRP	4.5	2.5	No	Centrifuge	No
Ne03	2013	gracilis	KRP	3	4	No	Centrifuge	No
Ne07	2013	rafflesiana	KRP	4.5	19	No	Centrifuge	Yes
Ne09	2013	gracilis	KRP	3	2.5	No	Centrifuge	Yes
Ne12	2013	ampullaria	KRP	3.5	11.2	No	Centrifuge	No
Ne13	2013	gracilis	KRP	3.5	2.9	No	Centrifuge	No
Ne16	2013	gracilis	KRP	2	1.3	No	Centrifuge	Yes
Ne18	2013	ampullaria	KRP	3.5	25	No	Centrifuge	Yes
Ne19	2013	gracilis	KRP	2	2	No	Centrifuge	Yes
Ne20	2013	rafflesiana	KRP	NA	27	No	Centrifuge	Yes
Ne21	2013	rafflesiana	KRP	NA	15	No	Centrifuge	Yes
Ne22	2013	rafflesiana	KRP	2	3.5	No	Centrifuge	Yes
Ne24	2013	rafflesiana	KRP	2.5	2.4	No	Centrifuge	Yes
Ne25	2013	rafflesiana	KRP	4	20	No	Centrifuge	Yes
Ne27	2013	ampullaria	KRP	4	4	No	Centrifuge	Yes
Ne28	2013	ampullaria	KRP	3.5	25	No	Centrifuge	Yes
Ne29	2013	ampullaria	KRP	4.5	15	No	Centrifuge	Yes

Supplementary Table 2.1 (Continued)

Ne30	2013	gracilis	KRP	2.5	2	Yes	Centrifuge	Yes
Ne31	2013	gracilis	KRP	1.5	0.9	Yes	Centrifuge	Yes
Ne32	2013	gracilis	KRP	2.5	3	Yes	Centrifuge	Yes
Ne33	2013	gracilis	BNTP	4.5	2	Yes	Centrifuge	Yes
Ne34	2013	rafflesiana	BNTP	6	20	No	Centrifuge	Yes
Ne36	2013	ampullaria	BNTP	5	5.1	No	Centrifuge	Yes
Ne37	2013	gracilis	BNTP	5	2.6	No	Centrifuge	Yes
Ne38	2013	gracilis	BNTP	5	2.3	Yes	Centrifuge	Yes
Ne39	2013	rafflesiana	BNTP	6.5	9.2	No	Centrifuge	Yes
Ne40	2013	ampullaria	BNTP	5	6.2	No	Centrifuge	Yes
Ne41	2013	ampullaria	BNTP	5	4.6	No	Centrifuge	Yes
Ne42	2013	rafflesiana	BNTP	5	0.5	No	Centrifuge	Yes
Ne44	2013	rafflesiana	BNTP	5	15	No	Centrifuge	Yes
Ne45	2013	ampullaria	BNTP	5	4	No	Centrifuge	Yes
Ne46	2013	rafflesiana	BNTP	6.5	2	No	Centrifuge	Yes
Ne47	2013	rafflesiana	BNTP	4.5	16	No	Centrifuge	Yes
Ne48	2013	ampullaria	BNTP	4.5	10.9	No	Centrifuge	Yes
Ne49	2013	gracilis	BNTP	2.5	6.4	No	Centrifuge	Yes
Ne50	2013	gracilis	BNTP	5.5	2	No	Centrifuge	Yes
Ne51	2013	ampullaria	BNTP	5	30	No	Centrifuge	Yes
Ne52	2013	gracilis	BNTP	6	3.5	No	Centrifuge	Yes
Ne53	2013	gracilis	BNTP	2	1.5	No	Centrifuge	Yes
Ne54	2013	ampullaria	BNTP	5	22.5	No	Centrifuge	Yes
Ne55	2013	ampullaria	BNTP	5	15.6	No	Centrifuge	Yes
Ne57	2013	rafflesiana	BNTP	7	6.5	No	Centrifuge	Yes
Ne58	2013	rafflesiana	UPR	NA	NA	No	Centrifuge	Yes
Ne59	2013	rafflesiana	UPR	2.5	7.5	No	Centrifuge	Yes
Ne60	2013	rafflesiana	UPR	5.5	7.6	No	Centrifuge	Yes
Ne62	2013	rafflesiana	UPR	4.5	40.2	No	Centrifuge	Yes
Ne63	2013	gracilis	UPR	2	1.75	Yes	Centrifuge	Yes
Ne64	2013	rafflesiana	UPR	5.5	3	No	Centrifuge	Yes
Ne65	2013	gracilis	UPR	3.5	2	No	Centrifuge	Yes
Ne66	2013	gracilis	UPR	2	1.5	No	Centrifuge	Yes
Ne67	2013	gracilis	UPR	2.5	4.4	Yes	Centrifuge	Yes
Ne68	2013	rafflesiana	UPR	5	15	No	Centrifuge	Yes
Ne70	2013	ampullaria	UPR	5	1.6	No	Centrifuge	Yes
Ne71	2013	ampullaria	UPR	5	0.1	No	Centrifuge	Yes
Ne72	2013	ampullaria	UPR	5	4	No	Centrifuge	Yes
Ne73	2013	rafflesiana	UPR	6.5	18	No	Centrifuge	Yes
Ne74	2013	rafflesiana	UPR	4.5	7.5	No	Centrifuge	Yes

Supplementary Table 2.2. The most abundant eukaryotic taxa from the *Nepenthes* samples

Taxon	% Of Total Seqs
Eukaryota; Opisthokonta; Metazoa; Arthropoda; Hexapoda; Insecta	41.90%
Eukaryota; Opisthokonta; Metazoa; Arthropoda; Chelicerata; Arachnida	8.10%
Eukaryota; Archaeplastida; Chloroplastida; Chlorophyta; Chlorophyceae; Pseudomuriella	6.40%
Eukaryota; SAR; Alveolata; Apicomplexa; Conoidasida; Gregarinasina	6.20%
Eukaryota; Cryptophyceae; Goniomonas; Goniomonas_sp._ATCC_50108;Other;Other	5.30%
Eukaryota; Opisthokonta; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes	4.30%
Eukaryota; SAR; Alveolata; Ciliophora; Intramacronucleata; Conthreep	3.40%
Eukaryota; Opisthokonta; Holozoa; Choanomonada; Craspedida; Lagenoeca	2.10%
Eukaryota; SAR; Rhizaria; Cercozoa; Glissomonadida; Heteromita	2.10%
Eukaryota; Opisthokonta; Metazoa; Annelida; Family_Incertae_Sedis; Aeolosoma	1.60%
Eukaryota; Excavata; Discoba; Discicristata; Euglenozoa; Euglenida	1.50%
Eukaryota; Opisthokonta; Metazoa; Rotifera; Philodinidae; uncultured_bdelloid_rotifer	1.50%
Eukaryota; Opisthokonta; Metazoa; Annelida; Family_Incertae_Sedis; Naididae	1.20%
Eukaryota; Archaeplastida; Chloroplastida; Chlorophyta; Trebouxiophyceae; Microthamnion	1.00%
Eukaryota; Excavata; Discoba; Discicristata; Euglenozoa; Kinetoplastea	0.80%
Eukaryota; SAR; Rhizaria; Cercozoa; Cercomonadidae; Cercomonas	0.80%
Eukaryota; Opisthokonta; Metazoa; Arthropoda; Crustacea; Malacostraca	0.70%
Eukaryota; SAR; Rhizaria; Cercozoa; Thecofilosea; Phaeodarea	0.70%
Eukaryota; Amoebozoa; Discosea; Longamoebia; Centramoebida; Acanthamoeba	0.50%
Other Taxa	9.90%

Supplementary Table 2.3. BLASTn taxonomy assignments of 18S and COI sequences

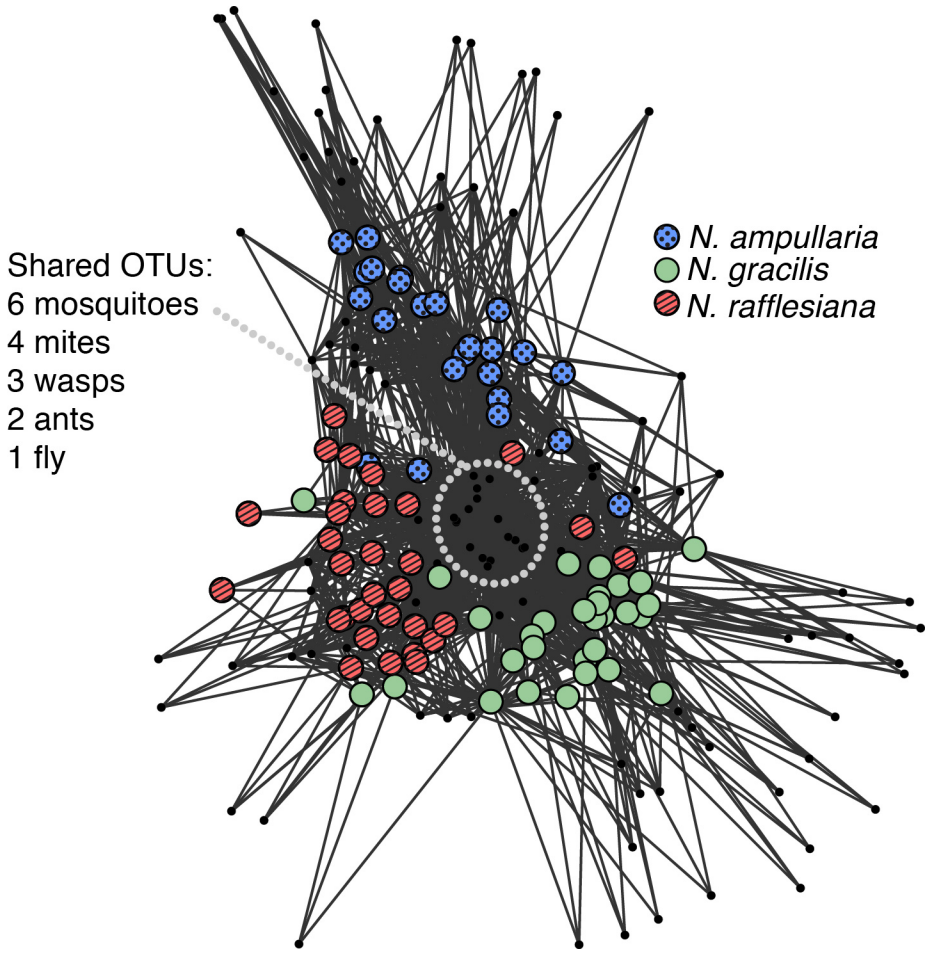
Original Name	BLAST Name	Top BLAST Matches	BLAST Family	Max score	Query cover	E-value	Identity
18S OTUs							
Culex_1_New.ReferenceOTU1293	Culex_A	Culex pilosus, C. peccator, C. mulrennani	Culicidae	261	100%	1E-66	98%
Culex_2_AAWU01030964	Culex_B	Culex pilosus, C. peccator, C. mulrennani	Culicidae	255	98%	6E-65	98%
Drosophila_1_New.ReferenceOTU490	Drosophila_A	Drosophila melanogaster, D. pseudoobscura, D. sechellia, D. persimilis	Drosophilidae	252	100%	8E-64	97%
Culex_3_New.ReferenceOTU306	Aedes_A	Aedes katherinensis, A. scutellaris, A. aegypti, A. albopictus, Drosophila willistoni	Culicidae	239	98%	6E-60	96%
Culex_4_New.ReferenceOTU1322	Culex_C	Culex lactator, C. thriambus, C. quinquefasciatus, C. interrogator, C. chidesteri, C. torrentium, C. tigripes, C. tarsalis	Culicidae	124	100%	2E-25	83%
Culex_5_New.ReferenceOTU115	Culex_D	Culex pilosus, C. peccator, C. mulrennani	Culicidae	250	98%	3E-63	97%
Culicoides_1_New.ReferenceOTU104	Culicoides_A	Culicoides imicola	Ceratopogonidae	219	100%	8E-54	93%
Drosophila_2_New.ReferenceOTU498	Drosophila_B	Drosophila melanogaster, D. pseudoobscura, D. sechellia, D. persimilis	Drosophilidae	246	100%	4E-62	96%
Corethrella_1_U49736	Corethrella_A	Corethrella wirthi	Corethrellidae	271	99%	1E-69	100%
Culex_6_New.ReferenceOTU267	Aedes_B	Aedes katherinensis, A. scutellaris, A. aegypti, A. albopictus, Drosophila willistoni	Culicidae	248	100%	1E-62	97%
Culex_7_New.ReferenceOTU567	Culex_E	Culex pilosus, C. peccator, C. mulrennani	Culicidae	246	96%	4E-62	98%
Oryza_1_New.ReferenceOTU640	Bradysia_A	Bradysia hygida	Sciaridae	237	100%	2E-59	95%
Culex_8_New.ReferenceOTU565	Culex_F	Culex pilosus, C. peccator, C. mulrennani	Culicidae	233	100%	3E-58	95%
Culex_9_New.ReferenceOTU644	Culex_G	Culex pilosus, C. peccator, C. mulrennani	Culicidae	217	100%	2E-53	96%
Toxorhynchites_1_New.ReferenceOTU739	Culex_H	Culex erythrothorax	Culicidae	179	100%	5E-42	85%
Culex_10_New.ReferenceOTU1211	Culex_I	Culex pilosus, C. peccator, C. mulrennani	Culicidae	230	100%	3E-57	96%
Culex_11_New.ReferenceOTU610	Culex_J	Culex pilosus, C. peccator, C. mulrennani	Culicidae	239	100%	6E-60	95%
Culex_12_New.ReferenceOTU606	Culex_K	Culex pilosus, C. peccator, C. mulrennani	Culicidae	241	99%	2E-60	96%
Culex_13_New.ReferenceOTU1196	Culex_L	Culex pilosus, C. peccator, C. mulrennani	Culicidae	239	100%	6E-60	95%
Oryza_2_New.ReferenceOTU1110	Bradysia_B	Bradysia hygida	Sciaridae	239	98%	6E-60	96%

Supplementary Table 2.3 (Continued)

Drosophila_3_New.ReferenceOTU81	Asilid_A	Ommatius sp.	Asilidae	235	97%	8E-59	95%
		Drosophila melanogaster, Lucilia porphyrina, Hemipyrellia ligurriens, Crysomya megacephala, D. pseudoobscura	Unclear, infraorder Muscomorpha	233	100%	3E-58	95%
Hermetia_1_EF650147	Dolichopodidae_A	Oncopygius distans	Dolichopodidae	199	100%	4E-48	98%
		Megaselia scalaris, Drosophila ananassae, Laxenecera albicincta	Unclear, infraorder Muscomorpha	194	100%	2E-46	97%
Culex_15_New.ReferenceOTU243	Aedes_C	Aedes katherinensis, A. scutellaris, A. aegypti, A. albopictus, Drosophila willistoni	Culicidae	259	100%	7E-66	98%
Culex_17_AAWU01013261	Culex_M	Culex pilosus, C. peccator, C. mulrennani	Culicidae	248	100%	1E-62	97%
COI barcodes Morphospecies (GenBank Accession)							
Endonepenthia_N24AE2 (KP845038)	Phoridae_1	Phoridae sp.	Phoridae	852	97%	0	90%
		Lucilia cf. illustris/caesar	Calliphoridae	836	100%	0	88%
Mosquito_N28AM (KP845039)	Tripteroides_1	Tripteroides bambusa	Culicidae	809	99%	0	89%
		Ochlerotatus camptorhynchus		805	99%	0	89%
Mosquito_N24AM (KP845040)	Ochlerotatus_1	Ochlerotatus camptorhynchus, Tripteroides bambusa	Culicidae	818	100%	0	89%
Lestodiplosis_N29AL (KP845041)	Cecidomyiidae_1	Cecidomyiidae sp.	Cecidomyiidae	818	98%	0	90%
Dasyhelea_N29AD (KP845042)	Lutzomyia_1	Lutzomyia longipalpis	Psychodidae	740	100%	0	86%
Corethrella_N38AC (KP845043)	Sergentomyia_1	Sergentomyia babu babu	Psychodidae	735	99%	0	86%
Endonepenthia_N18BEE (KP845044)	Phoridae_2	Phoridae sp.	Phoridae	809	97%	0	89%
		Lucilia caesar, L. cf. illustris/caesar, L. thatuna,	Calliphoridae	801	100%	0	89%
Mosquito_N22AM2 (KP845045)	Tripteroides_2	Tripteroides bambusa	Culicidae	814	99%	0	89%
Lestodiplosis_Ne39L (KP845046)	Phoridae_3	Phoridae sp.	Phoridae	829	97%	0	90%
		Megaselia scalaris	Phoridae	821	100%	0	89%
Endonepenthia_N22AE2 (KP845047)	Phoridae_4	Phoridae sp.	Phoridae	832	97%	0	90%
		Lucilia caesar, L. cf. illustris/caesar	Calliphoridae	816	99%	0	89%
Mosquito_Ne39M (KP845048)	Armigeres_1	Armigeres subalbatus	Culicidae	893	99%	0	91%

Supplementary Table 2.3 (Continued)

Corethrella_N38AC4 (KP845049)	Sergentomyia_2	Sergentomyia babu babu	Psychodidae	740	99%	0	86%
Dasyhelea_N19D (KP84505)	Drosophila_1	Drosophila neoguaramunu	Drosophilidae	695	99%	0	84%
		Verticia orientalis	Calliphoridae	684	100%	0	84%
		Culicoides immaculatus, Chromatomyia clematoides	Ceratopogonidae , Agromyzidae	682	100%	0	84%
Mosquito_N21AM3 (KP845051)	Ochlerotatus_2	Ochlerotatus camptorhynchus	Culicidae	810	100%	0	88%
Endonepenthia_N34AE (KP845052)	Phoridae_5	Phoridae sp.	Phoridae	825	98%	0	90%
		Lucilia cf. illustris/caesar	Calliphoridae	816	100%	0	89%
Mosquito_N36AM (KP845053)	Culex_1	Culex infantulus	Culicidae	908	98%	0	92%
Mosquito_N39BM3 (KP845054)	Ochlerotatus_3	Ochlerotatus camptorhynchus	Culicidae	814	100%	0	89%
Mosquito_N17AM (KP845055)	Ochlerotatus_4	Ochlerotatus camptorhynchus	Culicidae	854	99%	0	90%
Mosquito_N23BM (KP845056)	Ochlerotatus_5	Ochlerotatus melanimon	Culicidae	816	98%	0	89%
		Ochlerotatus camptorhynchus	Culicidae	816	100%	0	89%
Mosquito_N36AM2 (KP845057)	Culex_2	Culex infantulus	Culicidae	931	98%	0	93%
Mosquito_N38B.M2 (KP845058)	Culex_3	Culex infantulus	Culicidae	937	98%	0	93%
Mosquito_N23AM (KP845059)	Ochlerotatus_6	Ochlerotatus aurifer	Culicidae	794	98%	0	88%
		Ochlerotatus camptorhynchus	Culicidae	794	100%	0	88%
Dasyhelea_N37A.D2 (KP845060)	Phlebotomus_1	Phlebotomus alexandri	Psychodidae	747	100%	0	86%



Supplementary Figure 2.1:
Spring-embedded network of arthropod OTUs from *Nepenthes* pitchers. Large circles are samples and small black circles are OTUs. All OTUs present in at least 10% of the samples for each host species were used to create the network. The network is organized so samples sharing more OTUs cluster closer together. The OTUs in the center of the network are shared by samples from all three host species.

Methods for Supplementary Figure 2.1:

A "spring-embedded" network clusters samples and OTUs (two types of "nodes") using an algorithm that causes nodes to repel each other, while connections act as springs (Shannon *et al.* 2003). OTUs are connected via edges to samples, and the edge-weight is calculated using the number of sequences in the OTU found in each sample. Nodes are organized to minimize forces in the network, so that samples sharing more OTUs tend to cluster closer together, and samples sharing high-abundance OTUs tend to cluster more closely than samples sharing low-abundance OTUs. In QIIME, we first split the rarefied arthropod OTU table by host species, and then subsampled the resulting tables to include only those OTUs present in 10% of the samples for each host. The OTU tables were recombined, and used to make a network, visualized in Cytoscape using edge weights and the spring-embedded layout.

The spring-embedded network has OTUs represented by small black circles, and samples represented by large coloured circles. We chose to include only OTUs found in at least 10% of the samples from each host species, so that the OTUs in the network are less likely to be present purely by chance. The set of OTUs in this network overlap with those in Figure 4, but are not equivalent. The OTUs in the center of the network are shared among many samples and all host species, and can be thought of as the core set of organisms associated with our Singapore *Nepenthes* samples.

Reference for Supplementary Methods 2.1:

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.

Supplementary Material from Chapter 3:

Supplementary Table 3.1: Samples and Metadata

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
10Ng.14	2	18S	E	TS	10	14	pitcher	Nep.	gracilis	squat	NA	NA
10Ng.21	2	16S	E	TS	10	21	pitcher	Nep.	gracilis	squat	3	NA
10Nr.21	2	both	E	TS	10	21	pitcher	Nep.	raffle.	squat	1.5	NA
10Spb.21	2	both	E	TS	10	21	pitcher	Sar.	purp-b	squat	4	NA
10Spp.21	2	both	E	TS	10	21	pitcher	Sar.	purp-p	squat	5	NA
11Ng.o.21	2	both	E	TS	11	21	pitcher	Nep.	gracilis	squat	2.75	NA
11Ng.y.21	2	both	E	TS	11	21	pitcher	Nep.	gracilis	squat	3	NA
11Spb.21	2	both	E	TS	11	21	pitcher	Sar.	purp-b	squat	4.5	NA
11Spp.21	2	both	E	TS	11	21	pitcher	Sar.	purp-p	squat	5	NA
12Ng.21	2	both	E	TS	12	21	pitcher	Nep.	gracilis	squat	3	NA
12Spb.21	2	both	E	TS	12	21	pitcher	Sar.	purp-b	squat	4	NA
12Spp.14	2	both	E	TS	12	14	pitcher	Sar.	purp-p	squat	NA	NA
13Na.b.21	2	both	E	TS	13	21	pitcher	Nep.	ampull.	squat	4	NA
13Na.p.21	2	both	E	TS	13	21	pitcher	Nep.	ampull.	squat	4.5	NA
13Spb.21	2	both	E	TS	13	21	pitcher	Sar.	purp-b	squat	4	NA
14Na.21	2	both	E	TS	14	21	pitcher	Nep.	ampull.	squat	5.5	NA
14Ng.21	2	both	E	TS	14	21	pitcher	Nep.	gracilis	squat	3	NA
15Na.21	2	both	E	TS	15	21	pitcher	Nep.	ampull.	squat	4.5	NA
15Ng.21	2	both	E	TS	15	21	pitcher	Nep.	gracilis	squat	3.5	NA
15Spb.21	2	both	E	TS	15	21	pitcher	Sar.	purp-b	squat	4.5	NA
16Na.21	2	both	E	TS	16	21	pitcher	Nep.	ampull.	squat	3	NA
16Ng.21	2	both	E	TS	16	21	pitcher	Nep.	gracilis	squat	7	NA
16Spb.b.21	2	both	E	TS	16	21	pitcher	Sar.	purp-b	squat	5	NA
16Spb.y.14	2	18S	E	TS	16	14	pitcher	Sar.	purp-b	squat	NA	NA
17Bog.49	2	both	E	TS	17	49	envmt	Bog	bog	NA	5	NA
17Gtp.49	2	18S	E	TS	17	49	tube	Tube	gtube-p	squat	5	NA
17Na.49	2	both	E	TS	17	49	pitcher	Nep.	ampull.	squat	5	NA
17Spb.49	2	both	E	TS	17	49	pitcher	Sar.	purp-b	squat	4.75	NA
17Spp.49	2	both	E	TS	17	49	pitcher	Sar.	purp-p	squat	5	NA
18Gt.49	2	both	E	TS	18	49	tube	Tube	gtube	squat	5	NA
18Gtp.49	2	both	E	TS	18	49	tube	Tube	gtube-p	squat	7.5	NA
18Na.49	2	both	E	TS	18	49	pitcher	Nep.	ampull.	squat	5	NA
18Spb.49	2	both	E	TS	18	49	pitcher	Sar.	purp-b	squat	4.75	NA
18Spp.49	2	both	E	TS	19	49	pitcher	Sar.	purp-p	squat	5	NA
19Gt.14	2	16S	E	TS	19	14	tube	Tube	gtube	squat	5	NA
19Gtp.49	2	both	E	TS	19	49	tube	Tube	gtube-p	squat	5	NA
19Na.49	2	both	E	TS	19	49	pitcher	Nep.	ampull.	squat	5	NA
19Spb.w.49	2	16S	E	TS	19	49	pitcher	Sar.	purp-b	squat	4.75	NA
19Spb.y.49	2	both	E	TS	19	49	pitcher	Sar.	purp-b	squat	5	NA
19Spp.49	2	both	E	TS	19	49	pitcher	Sar.	purp-p	squat	4.75	NA
20Gtp.49	2	both	E	TS	20	49	tube	Tube	gtube-p	squat	5	NA
20Na.49	2	both	E	TS	20	49	pitcher	Nep.	ampull.	squat	5	NA
20Spb.y.49	2	both	E	TS	20	49	pitcher	Sar.	purp-b	squat	4.5	NA
20Spp.b.49	2	both	E	TS	20	49	pitcher	Sar.	purp-p	squat	4.75	NA
21Gtp.49	2	both	E	TS	21	49	tube	Tube	gtube-p	squat	5	NA
21Na.o.49	2	both	E	TS	21	49	pitcher	Nep.	ampull.	squat	5	NA
21Na.y.49	2	both	E	TS	21	49	pitcher	Nep.	ampull.	squat	5	NA

Supplementary Table 3.1 (Continued)

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
21Nb.49	2	both	E	TS	21	49	pitcher	Nep.	bicalc.	squat	5	NA
21Spb.p.49	2	both	E	TS	21	49	pitcher	Sar.	purp-b	squat	5	NA
21Spp.49	2	both	E	TS	21	49	pitcher	Sar.	purp-p	squat	4.75	NA
22Bog.48	2	both	E	TS	22	48	envmt	Bog	bog	NA	4.5	NA
22Gtp.48	2	both	E	TS	22	48	tube	Tube	gtube-p	squat	5	NA
22Na.35	2	both	E	TS	22	35	pitcher	Nep.	ampull.	squat	4.75	NA
22Nb.35	2	both	E	TS	22	35	pitcher	Nep.	bicalc.	squat	4.75	NA
22Nr.48	2	both	E	TS	22	48	pitcher	Nep.	raffle.	squat	4.25	NA
22Spb.48	2	both	E	TS	22	48	pitcher	Sar.	purp-b	squat	4.75	NA
23Bog.48	2	18S	E	TS	23	48	envmt	Bog	bog	NA	4.5	NA
23Gt.48	2	both	E	TS	23	48	tube	Tube	gtube	squat	4.75	NA
23Gtp.48	2	both	E	TS	23	48	tube	Tube	gtube-p	squat	4.75	NA
23Na.15	2	both	E	TS	23	15	pitcher	Nep.	ampull.	squat	5.25	NA
23Nb.15	2	both	E	TS	23	15	pitcher	Nep.	bicalc.	squat	4.75	NA
23Nr.35	2	both	E	TS	23	35	pitcher	Nep.	raffle.	squat	2.75	NA
23Spb.48	2	both	E	TS	23	48	pitcher	Sar.	purp-b	squat	4.75	NA
24Bog.48	2	18S	E	TS	24	48	envmt	Bog	bog	NA	4.5	NA
24Gt.48	2	both	E	TS	24	48	tube	Tube	gtube	squat	5	NA
24Gtp.48	2	both	E	TS	24	48	tube	Tube	gtube-p	squat	4.75	NA
24Na.48	2	both	E	TS	24	48	pitcher	Nep.	ampull.	squat	7	NA
24Nb.48	2	both	E	TS	24	48	pitcher	Nep.	bicalc.	squat	4.75	NA
24Nr.48	2	both	E	TS	24	48	pitcher	Nep.	raffle.	squat	3.5	NA
24Spb.48	2	both	E	TS	24	48	pitcher	Sar.	purp-b	squat	6.5	NA
25Bog.48	2	both	E	TS	25	48	envmt	Bog	bog	NA	4.5	NA
25Gt.48	2	both	E	TS	25	48	tube	Tube	gtube	squat	5.5	NA
25Gtp.48	2	both	E	TS	25	48	tube	Tube	gtube-p	squat	5	NA
25Na.r.48	2	both	E	TS	25	48	pitcher	Nep.	ampull.	squat	4.75	NA
25Na.y.48	2	both	E	TS	25	48	pitcher	Nep.	ampull.	squat	4.75	NA
25Nb.48	2	both	E	TS	25	48	pitcher	Nep.	bicalc.	squat	4.75	NA
25Spb.48	2	both	E	TS	25	48	pitcher	Sar.	purp-b	squat	4.75	NA
26Bog.48	2	both	E	TS	26	48	envmt	Bog	bog	NA	4.75	NA
26Gt.48	2	16S	E	TS	26	48	tube	Tube	gtube	squat	4.75	NA
26Gtp.48	2	both	E	TS	26	48	tube	Tube	gtube-p	squat	4.75	NA
26Na.r.48	2	both	E	TS	26	48	pitcher	Nep.	ampull.	squat	4.75	NA
26Na.y.48	2	both	E	TS	26	48	pitcher	Nep.	ampull.	squat	6.5	NA
26Nb.15	2	both	E	TS	26	15	pitcher	Nep.	bicalc.	squat	4.75	NA
26Spb.48	2	both	E	TS	26	48	pitcher	Sar.	purp-b	squat	6.5	NA
9Ng.21	2	both	E	TS	9	21	pitcher	Nep.	gracilis	squat	3	NA
9Nr.21	2	16S	E	TS	9	21	pitcher	Nep.	raffle.	squat	2	NA
9Spp.21	2	both	E	TS	9	21	pitcher	Sar.	purp-p	squat	6	NA
Bog01	1	both	C	Sing	BTNP	NA	envmt	Bog	bog	NA	NA	NA
BogBF01	2	18S	C	GC	BF	NA	envmt	Bog	bog	NA	5.25	7.5
BogBF03	2	both	C	GC	BF	NA	envmt	Bog	bog	NA	5	7.2
BogBF04	2	16S	C	GC	BF	NA	envmt	Bog	bog	NA	5.25	7.6
BogBW02	2	16S	C	GC	BW	NA	envmt	Bog	bog	NA	4.75	7.5
BogBW03	2	both	C	GC	BW	NA	envmt	Bog	bog	NA	4.75	7
BogBW04	2	both	C	GC	BW	NA	envmt	Bog	bog	NA	5	7.2
BogBW05	2	both	C	GC	BW	NA	envmt	Bog	bog	NA	4.75	7.5
BogEB02	2	both	C	GC	EB1	NA	envmt	Bog	bog	NA	5	7.2
BogEB04	2	18S	C	GC	EB1	NA	envmt	Bog	bog	NA	4.75	7.5
BogNP01	2	both	C	GC	NP	NA	envmt	Bog	bog	NA	5	7.5

Supplementary Table 3.1 (Continued)

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
BogNP02	2	both	C	GC	NP	NA	envmt	Bog	bog	NA	5	7.5
BogOFB01	2	both	C	GC	OFB	NA	envmt	Bog	bog	NA	5	7.5
BogSHB01	2	16S	C	GC	SH	NA	envmt	Bog	bog	NA	4.75	7.3
BogSHB02	2	both	C	GC	SH	NA	envmt	Bog	bog	NA	5	7.7
BogSHB03	2	both	C	GC	SH	NA	envmt	Bog	bog	NA	4.75	7.5
BogSHB04	2	both	C	GC	SH	NA	envmt	Bog	bog	NA	4.5	7.3
BogSHB05	2	both	C	GC	SH	NA	envmt	Bog	bog	NA	4.75	7.4
BogSHB06	2	both	C	GC	SH	NA	envmt	Bog	bog	NA	4.75	7.6
BogSHB07	2	both	C	GC	SH	NA	envmt	Bog	bog	NA	4.75	7
BogTS03	2	18S	C	HF	TS	NA	envmt	Bog	bog	NA	4.5	7
BogWB01	2	both	C	GC	WB	NA	envmt	Bog	bog	NA	5	7.7
L01	2	both	C	MB	NG	NA	envmt	Leaf	leaf	NA	4.5	4
L02	2	both	C	MB	NG	NA	envmt	Leaf	leaf	NA	4.5	4.5
L03	2	both	C	MB	NG	NA	envmt	Leaf	leaf	NA	4.75	3.3
L04	2	both	C	MB	NG	NA	envmt	Leaf	leaf	NA	4.75	0.9
L05	2	16S	C	MB	NG	NA	envmt	Leaf	leaf	NA	5	13
L06	2	18S	C	MB	NG	NA	envmt	Leaf	leaf	NA	4.75	1.8
L07	2	both	C	MB	NG	NA	envmt	Leaf	leaf	NA	4.75	3.1
L09	2	both	C	MB	NG	NA	envmt	Leaf	leaf	NA	4.75	2
L10	2	both	C	MB	NG	NA	envmt	Leaf	leaf	NA	4.75	2.4
Lw03	1	both	C	Sing	BTNP	NA	envmt	Leaf	leaf	NA	NA	NA
MB01	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	2	0.9
MB03	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	5	40
MB04	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	4.75	45
MB05	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	3.5	18
MB06	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	4	500
MB07	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	5	2.9
MB08	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	4.75	8.2
MB09	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	4.75	5
MB10	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	5	5.6
MB11	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	4.5	6.4
MB12	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	4.5	13.3
MB13	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	3.5	6.5
MB15	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	2	11.6
MB16	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	2	5.8
MB18	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	1.75	7.3
MB19	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	4.75	100
MB20	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	2	25
MB21	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	4.75	20
MB22	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	3.5	150
MB23	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	5	5.4
MB24	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	4.75	150
MB25	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	2	22
MB26	2	16S	C	MB	NG	NA	pitcher	Nep.	steno.	squat	4.75	18
MB27	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	1.75	20
MB28	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	7	26
MB29	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	5	19
MB30	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	4	34
MB33	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	2	28
MB34	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	1.5	1.3
MB35	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	2	4

Supplementary Table 3.1 (Continued)

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
MB36	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	4.5	70
MB37	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	4.5	500
MB38	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	3.75	55
MB39	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	1.5	45
MB40	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	3	75
MB42	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	5	120
MB44	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	4.25	100
MB45	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	5.25	5.5
MB46	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	1.5	9.6
MB49	2	both	C	MB	NG	NA	pitcher	Nep.	veitchii	squat	3.5	500
MB50	2	both	C	MB	NG	NA	pitcher	Nep.	tentac.	squat	5	7
MB51	2	both	C	MB	NG	NA	pitcher	Nep.	tentac.	squat	4.75	5.2
MB52	2	both	C	MB	NG	NA	pitcher	Nep.	tentac.	squat	5	1.7
MB53	2	both	C	MB	NG	NA	pitcher	Nep.	tentac.	squat	7.5	1.4
MB54	2	both	C	MB	NG	NA	pitcher	Nep.	tentac.	squat	7.5	0.7
MB55	2	both	C	MB	NG	NA	pitcher	Nep.	tentac.	squat	4.5	3.1
MB56	2	both	C	MB	NG	NA	pitcher	Nep.	tentac.	squat	5	1.5
MB57	2	both	C	MB	NG	NA	pitcher	Nep.	tentac.	squat	4	1.7
MB58	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	2	18
MB59	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	5	8.1
MB60	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	1.75	2.6
MB61	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5	8
MB62	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5	8
MB63	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5	6.8
MB64	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5	4.9
MB65	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5	2.6
MB66	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5	4.9
MB67	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	4.75	2.5
MB68	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5	2.1
MB69	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5.5	6.1
MB70	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	4.75	8.2
MB71	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	4.75	1.5
MB72	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5	8.9
MB73	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5	1.5
MB74	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	4.5	12.5
MB75	2	16S	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	7	2.5
MB76	2	both	C	MB	NG	NA	pitcher	Nep.	hirsuta	squat	5	4.3
MB77	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	5	5.7
MB78	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	5	5.2
MB79	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	5	2
MB80	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	5	8.4
MB81	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	4.75	5.7
MB82	2	16S	C	MB	NG	NA	pitcher	Nep.	reint.	squat	4	8.9
MB83	2	both	C	MB	NG	NA	pitcher	Nep.	reint.	squat	5	10
MB84	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	5.5	25
MB85	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	1.75	10.6
MB86	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	2	4.9
MB87	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	5	35
MB88	2	both	C	MB	NG	NA	pitcher	Nep.	steno.	squat	1.5	1.4
N01A	1	16S	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N01B	1	both	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	NA	NA

Supplementary Table 3.1 (Continued)

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
N02A	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	NA	NA
N02B	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	NA	NA
N02C	1	16S	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	NA	NA
N03C	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	NA	NA
N05B	1	16S	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N06A	1	16S	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N06B	1	16S	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N06C	1	both	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N07A	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	NA	NA
N07B	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	NA	NA
N08A	1	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	NA	NA
N10A	1	16S	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	NA	NA
N10B.	1	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	NA	NA
N12	1	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	NA	NA
N13	1	both	C	Sing	UPR	NA	pitcher	Nep.	gracilis	squat	NA	NA
N15	1	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	NA	NA
N20	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N21	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N22	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N23A	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N23B	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N24	1	18S	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N27A	1	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	NA	NA
N27C	1	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	NA	NA
N28	1	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	NA	NA
N29A	1	18S	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	NA	NA
N32	1	18S	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	NA	NA
N33	1	18S	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	NA	NA
N34	1	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	NA	NA
N37A.	1	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	NA	NA
N37B	1	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	NA	NA
N38B.	1	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	NA	NA
NaB04	2	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	2
NaB05	2	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	1.9
NaB06	2	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	4.75	4.7
NaB07	2	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	5.1
NaB08	2	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	12.4
NaB10	2	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	1.8
NaK26	2	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	5	6.3
NaK27	2	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	5	0.7
NaK31	2	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	4.75	1.2
NaK32	2	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	4.5	0.9
NaU23	2	both	C	Sing	UPR	NA	pitcher	Nep.	ampull.	squat	4.5	2.4
Ne02	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	4.5	2.5
Ne03	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	3	4
Ne07	1	18S	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	4.5	19
Ne09	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	3	2.5
Ne12	1	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	3.5	11.2
Ne13	1	18S	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	3.5	2.9
Ne16	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	2	1.3
Ne18	1	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	3.5	25

Supplementary Table 3.1 (Continued)

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
Ne19	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	2	2
Ne20	1	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	NA	27
Ne21	1	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	NA	15
Ne22	1	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	2	3.5
Ne24	1	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	2.5	2.4
Ne25	1	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	4	20
Ne27	1	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	4	4
Ne28	1	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	3.5	25
Ne29	1	both	C	Sing	KRP	NA	pitcher	Nep.	ampull.	squat	4.5	15
Ne30	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	2.5	2
Ne31	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	1.5	0.9
Ne32	1	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	2.5	3
Ne33	1	both	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	4.5	2
Ne34	1	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	6	20
Ne36	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	5.1
Ne37	1	both	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	5	2.6
Ne38	1	both	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	5	2.3
Ne39	1	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	6.5	9.2
Ne40	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	6.2
Ne41	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	4.6
Ne42	1	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	5	0.5
Ne44	1	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	5	15
Ne45	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	4
Ne46	1	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	6.5	2
Ne47	1	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	4.5	16
Ne48	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	4.5	10.9
Ne49	1	both	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	2.5	6.4
Ne50	1	both	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	5.5	2
Ne51	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	30
Ne52	1	both	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	6	3.5
Ne53	1	both	C	Sing	BTNP	NA	pitcher	Nep.	gracilis	squat	2	1.5
Ne54	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	22.5
Ne55	1	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	5	15.6
Ne57	1	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	7	6.5
Ne58	1	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	NA	NA
Ne59	1	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	2.5	7.5
Ne60	1	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	5.5	7.6
Ne62	1	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	4.5	40.2
Ne63	1	both	C	Sing	UPR	NA	pitcher	Nep.	gracilis	squat	2	1.75
Ne64	1	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	5.5	3
Ne65	1	both	C	Sing	UPR	NA	pitcher	Nep.	gracilis	squat	3.5	2
Ne66	1	both	C	Sing	UPR	NA	pitcher	Nep.	gracilis	squat	2	1.5
Ne67	1	both	C	Sing	UPR	NA	pitcher	Nep.	gracilis	squat	2.5	4.4
Ne68	1	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	5	15
Ne70	1	both	C	Sing	UPR	NA	pitcher	Nep.	ampull.	squat	5	1.6
Ne71	1	both	C	Sing	UPR	NA	pitcher	Nep.	ampull.	squat	5	0.1
Ne72	1	18S	C	Sing	UPR	NA	pitcher	Nep.	ampull.	squat	5	4
Ne73	1	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	6.5	18
Ne74	1	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	4.5	7.5
NrB01	2	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	4.5	7.6
NrB02	2	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	4.75	1.2

Supplementary Table 3.1 (Continued)

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
NrB03	2	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	4.75	6.1
NrB09	2	both	C	Sing	BTNP	NA	pitcher	Nep.	raffle.	squat	4.5	1
NrK29	2	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	1.75	9.2
NrK30	2	both	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	2	8.6
NrK34	2	18S	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	5	0.5
NrU16	2	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	4	0.2
NrU17	2	both	C	Sing	UPR	NA	pitcher	Nep.	raffle.	squat	5	12.5
Nuna02	2	both	C	Sing	BTNP	NA	pitcher	Nep.	ampull.	squat	4.5	0.2
Nung08	2	both	C	Sing	KRP	NA	pitcher	Nep.	gracilis	squat	2	0.01
Nunr12	2	18S	C	Sing	KRP	NA	pitcher	Nep.	raffle.	squat	4	0.1
S01	2	both	C	MB	NG	NA	envmt	Soil	soil	NA	NA	NA
S02	2	both	C	MB	NG	NA	envmt	Soil	soil	NA	NA	NA
S03	2	both	C	MB	NG	NA	envmt	Soil	soil	NA	NA	NA
S04	2	16S	C	MB	NG	NA	envmt	Soil	soil	NA	NA	NA
S05	2	16S	C	MB	NG	NA	envmt	Soil	soil	NA	NA	NA
S06	2	both	C	MB	NG	NA	envmt	Soil	soil	NA	NA	NA
S07	2	both	C	MB	NG	NA	envmt	Soil	soil	NA	NA	NA
S08	2	both	C	MB	NG	NA	envmt	Soil	soil	NA	NA	NA
S09	2	16S	C	MB	NG	NA	envmt	Soil	soil	NA	NA	NA
S10	2	both	C	MB	NG	NA	envmt	Soil	soil	NA	NA	NA
Sa01	2	both	C	GC	SB	NA	pitcher	Sar.	alata	tall	7.5	0.6
Sa02	2	both	C	GC	SB	NA	pitcher	Sar.	alata	tall	7.5	0.5
Sa03	2	16S	C	GC	SB	NA	pitcher	Sar.	alata	tall	5.75	0.7
Sa05	2	both	C	GC	SB	NA	pitcher	Sar.	alata	tall	5	2.2
Sa06	2	both	C	GC	SB	NA	pitcher	Sar.	alata	tall	5.25	1
Sa07	2	both	C	GC	SB	NA	pitcher	Sar.	alata	tall	4.75	3.9
Sa08	2	16S	C	GC	SB	NA	pitcher	Sar.	alata	tall	7	0.5
Sa09	2	both	C	GC	BF	NA	pitcher	Sar.	alata	tall	7	0.4
Sa14	2	16S	C	GC	BF	NA	pitcher	Sar.	alata	tall	4	0.5
Sa15	2	both	C	GC	BF	NA	pitcher	Sar.	alata	tall	4.75	4.5
Sa16	2	both	C	GC	BF	NA	pitcher	Sar.	alata	tall	4	1
Sa17	2	both	C	GC	BF	NA	pitcher	Sar.	alata	tall	5	0.1
Sa18	2	both	C	GC	BF	NA	pitcher	Sar.	alata	tall	4	2.8
Sa19	2	both	C	GC	OFB	NA	pitcher	Sar.	alata	tall	4.25	1
Sa20	2	both	C	GC	OFB	NA	pitcher	Sar.	alata	tall	4	6.4
Sa21	2	both	C	GC	OFB	NA	pitcher	Sar.	alata	tall	4.5	7.8
Sa22	2	both	C	GC	OFB	NA	pitcher	Sar.	alata	tall	4.75	3.2
Sa23	2	both	C	GC	OFB	NA	pitcher	Sar.	alata	tall	3	2.9
Sa24	2	both	C	GC	OFB	NA	pitcher	Sar.	alata	tall	4.5	1.4
Sa25	2	both	C	GC	OFB	NA	pitcher	Sar.	alata	tall	4.5	1.7
Sa26	2	both	C	GC	OFB	NA	pitcher	Sar.	alata	tall	4.25	6.9
Sa27	2	both	C	GC	WB	NA	pitcher	Sar.	alata	tall	4.75	1.5
Sa28	2	both	C	GC	WB	NA	pitcher	Sar.	alata	tall	4	5.8
Sa29	2	both	C	GC	WB	NA	pitcher	Sar.	alata	tall	5.5	8.1
Sf01	2	both	C	GC	BW	NA	pitcher	Sar.	flava	tall	5	11.5
Sf02	2	both	C	GC	BW	NA	pitcher	Sar.	flava	tall	4	4.2
Sf03	2	both	C	GC	BW	NA	pitcher	Sar.	flava	tall	4.75	18
Sf04	2	both	C	GC	BW	NA	pitcher	Sar.	flava	tall	4.75	8.2
Sf05	2	both	C	GC	BW	NA	pitcher	Sar.	flava	tall	4.75	30
Sf06	2	both	C	GC	BW	NA	pitcher	Sar.	flava	tall	4.75	13
Sf07	2	both	C	GC	BW	NA	pitcher	Sar.	flava	tall	5	12.3

Supplementary Table 3.1 (Continued)

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
Sf08	2	both	C	GC	BW	NA	pitcher	Sar.	flava	tall	6.5	8.5
Sf09	2	16S	C	GC	BW	NA	pitcher	Sar.	flava	tall	6	3.6
Sf11	2	both	C	GC	BW	NA	pitcher	Sar.	flava	tall	5	2.8
Sf12	2	both	C	GC	NP	NA	pitcher	Sar.	flava	tall	4	4.4
Sf14	2	both	C	GC	NP	NA	pitcher	Sar.	flava	tall	4	2.9
Sf16	2	16S	C	GC	NP	NA	pitcher	Sar.	flava	tall	4.25	3.4
Sf18	2	both	C	GC	MS	NA	pitcher	Sar.	flava	tall	7	2.8
Sf21	2	16S	C	GC	MS	NA	pitcher	Sar.	flava	tall	7	2.5
Sf22	2	both	C	GC	MS	NA	pitcher	Sar.	flava	tall	5	5
Sf23	2	16S	C	GC	MS	NA	pitcher	Sar.	flava	tall	6.5	4
Sf24	2	both	C	GC	CB	NA	pitcher	Sar.	flava	tall	5.25	7.5
Sf25	2	both	C	GC	CB	NA	pitcher	Sar.	flava	tall	4	4.2
Sf27	2	both	C	GC	CB	NA	pitcher	Sar.	flava	tall	6.5	2.9
Sf28	2	both	C	GC	CB	NA	pitcher	Sar.	flava	tall	3.75	3.4
Sf29	2	both	C	GC	CB	NA	pitcher	Sar.	flava	tall	4	1.5
Sf30	2	both	C	GC	PPS	NA	pitcher	Sar.	flava	tall	4.75	3.7
Sf31	2	both	C	GC	PPS	NA	pitcher	Sar.	flava	tall	4	4.2
Sf32	2	both	C	GC	PPS	NA	pitcher	Sar.	flava	tall	4	2.3
Sf33	2	both	C	GC	PPS	NA	pitcher	Sar.	flava	tall	5	6.5
Sf34	2	16S	C	GC	PPS	NA	pitcher	Sar.	flava	tall	5	3
Sf37	2	16S	C	GC	EB1	NA	pitcher	Sar.	flava	tall	6	1.6
Sf38	2	both	C	GC	EB1	NA	pitcher	Sar.	flava	tall	7	1.5
Sf41	2	16S	C	GC	EB1	NA	pitcher	Sar.	flava	tall	5	1.8
Sl01	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	7	4
Sl02	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	6.5	0.3
Sl06	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	4.75	12.5
Sl07	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	5	4
Sl08	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	5	5.1
Sl10	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	5	2.5
Sl11	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	5	1.9
Sl13	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	7	6.4
Sl14	2	16S	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	8	0.5
Sl15	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	4.75	2.1
Sl16	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	3.5	1.8
Sl17	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	5	2.6
Sl18	2	both	C	GC	SH	NA	pitcher	Sar.	leuco.	tall	5	0.4
Sl19	2	both	C	GC	BW	NA	pitcher	Sar.	leuco.	tall	4.5	6.5
Sl20	2	both	C	GC	BW	NA	pitcher	Sar.	leuco.	tall	4.75	20
Sl21	2	both	C	GC	BW	NA	pitcher	Sar.	leuco.	tall	4.75	4.8
Sl22	2	both	C	GC	BW	NA	pitcher	Sar.	leuco.	tall	4.75	7.5
Sl23	2	both	C	GC	BW	NA	pitcher	Sar.	leuco.	tall	4.75	27
Sl24	2	both	C	GC	BW	NA	pitcher	Sar.	leuco.	tall	4.75	13.2
Sl25	2	both	C	GC	NB	NA	pitcher	Sar.	leuco.	tall	6.5	6
Sl26	2	16S	C	GC	NB	NA	pitcher	Sar.	leuco.	tall	5	2.8
Sl27	2	both	C	GC	NB	NA	pitcher	Sar.	leuco.	tall	5.5	3.6
Sl28	2	both	C	GC	NB	NA	pitcher	Sar.	leuco.	tall	4.75	15
Sl29	2	both	C	GC	NB	NA	pitcher	Sar.	leuco.	tall	5	30
Sl30	2	16S	C	GC	NB	NA	pitcher	Sar.	leuco.	tall	5	2.7
Sl31	2	both	C	GC	EB1	NA	pitcher	Sar.	leuco.	tall	5	8.6
Sl32	2	16S	C	GC	EB1	NA	pitcher	Sar.	leuco.	tall	5.25	4.5
Sl33	2	both	C	GC	EB1	NA	pitcher	Sar.	leuco.	tall	5.5	4.8

Supplementary Table 3.1 (Continued)

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
Sl34	2	16S	C	GC	EB2	NA	pitcher	Sar.	leuco.	tall	8.5	0.2
Sl36	2	both	C	GC	EB2	NA	pitcher	Sar.	leuco.	tall	8.5	0.4
Sl37	2	16S	C	GC	EB2	NA	pitcher	Sar.	leuco.	tall	7	2
Sl38	2	16S	C	GC	EB2	NA	pitcher	Sar.	leuco.	tall	8	0.5
Sl39	2	both	C	GC	WB	NA	pitcher	Sar.	leuco.	tall	5	13
Sl40	2	both	C	GC	WB	NA	pitcher	Sar.	leuco.	tall	4.75	22
Sl41	2	both	C	GC	WB	NA	pitcher	Sar.	leuco.	tall	5	6.9
SoilPPS01	2	both	C	GC	PPS	NA	envmt	Soil	soil	NA	NA	NA
SoilPPS03	2	both	C	GC	PPS	NA	envmt	Soil	soil	NA	NA	NA
SoilSH01	2	both	C	GC	SH	NA	envmt	Soil	soil	NA	NA	NA
Sp01	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	5	9.4
Sp02	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	5	9.7
Sp03	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.75	10.9
Sp04	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.25	14
Sp05	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.5	4
Sp06	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.5	43
Sp07	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4	6
Sp08	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.75	10.8
Sp09	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	3.5	3.5
Sp10	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.75	10.5
Sp11	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.75	3
Sp12	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.75	7.6
Sp13	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.5	4.5
Sp14	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.5	18
Sp15	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.75	9.8
Sp16	2	18S	C	HF	TS	NA	pitcher	Sar.	purp	squat	7	4.1
Sp17	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	5	16
Sp18	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.5	11.9
Sp19	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	7	4.7
Sp20	2	both	C	HF	TS	NA	pitcher	Sar.	purp	squat	4.25	20
Sro02	2	both	C	GC	SH	NA	pitcher	Sar.	rosea	squat	5	20
Sro04	2	both	C	GC	SH	NA	pitcher	Sar.	rosea	squat	4.5	24
Sro07	2	both	C	GC	SH	NA	pitcher	Sar.	rosea	squat	4.75	13
Sro08	2	both	C	GC	SH	NA	pitcher	Sar.	rosea	squat	7	6
Sro09	2	both	C	GC	SH	NA	pitcher	Sar.	rosea	squat	5	5.1
Sro10	2	both	C	GC	SH	NA	pitcher	Sar.	rosea	squat	7	11
Sro11	2	both	C	GC	SH	NA	pitcher	Sar.	rosea	squat	4.5	22
Sro12	2	both	C	GC	SH	NA	pitcher	Sar.	rosea	squat	6.5	17
Sro14	2	both	C	GC	MS	NA	pitcher	Sar.	rosea	squat	5.25	6
Sro15	2	both	C	GC	MS	NA	pitcher	Sar.	rosea	squat	5	4.8
Sro16	2	both	C	GC	MS	NA	pitcher	Sar.	rosea	squat	7	9.4
Sro17	2	both	C	GC	MS	NA	pitcher	Sar.	rosea	squat	5	8.5
Sro18	2	both	C	GC	MS	NA	pitcher	Sar.	rosea	squat	4.5	30
Sro20	2	both	C	GC	CB	NA	pitcher	Sar.	rosea	squat	4.25	7
Sro21	2	both	C	GC	CB	NA	pitcher	Sar.	rosea	squat	7	5.4
Sro22	2	both	C	GC	CB	NA	pitcher	Sar.	rosea	squat	5	7
Sro23	2	both	C	GC	CB	NA	pitcher	Sar.	rosea	squat	5	12
Sro24	2	both	C	GC	CB	NA	pitcher	Sar.	rosea	squat	4.25	27
Sro25	2	both	C	GC	PPS	NA	pitcher	Sar.	rosea	squat	5.5	10.1
Sro26	2	both	C	GC	PPS	NA	pitcher	Sar.	rosea	squat	7	11.3
Sro27	2	both	C	GC	PPS	NA	pitcher	Sar.	rosea	squat	7.5	12.3

Supplementary Table 3.1 (Continued)

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
Sro28	2	both	C	GC	PPS	NA	pitcher	Sar.	rosea	squat	7.5	27
Sro30	2	both	C	GC	PPS	NA	pitcher	Sar.	rosea	squat	5.5	8
Sru01	2	both	C	GC	SH	NA	pitcher	Sar.	rubra	tall	5	0.5
Sru02	2	both	C	GC	SH	NA	pitcher	Sar.	rubra	tall	4.75	2
Sru03	2	both	C	GC	SH	NA	pitcher	Sar.	rubra	tall	7	0.4
Sru04	2	both	C	GC	SH	NA	pitcher	Sar.	rubra	tall	5	12.4
Sru05	2	both	C	GC	SH	NA	pitcher	Sar.	rubra	tall	5.25	0.6
Sru06	2	both	C	GC	SH	NA	pitcher	Sar.	rubra	tall	4.75	1.4
Sru08	2	both	C	GC	EB1	NA	pitcher	Sar.	rubra	tall	5	0.2
Sru09	2	16S	C	GC	EB1	NA	pitcher	Sar.	rubra	tall	7	3.9
Sru10	2	16S	C	GC	EB1	NA	pitcher	Sar.	rubra	tall	7	2.9
Sru12	2	both	C	GC	EB1	NA	pitcher	Sar.	rubra	tall	8	0.8
Sru14	2	both	C	GC	EB2	NA	pitcher	Sar.	rubra	tall	8.5	0.3
T1M1	1	18S	C	Sing	KRP	NA	envmt	Soil	soil	NA	NA	NA
T1M2	1	18S	C	Sing	KRP	NA	envmt	Soil	soil	NA	NA	NA
T1M3	1	both	C	Sing	KRP	NA	envmt	Soil	soil	NA	NA	NA
T1M4	1	both	C	Sing	KRP	NA	envmt	Soil	soil	NA	NA	NA
T1M5	1	both	C	Sing	KRP	NA	envmt	Soil	soil	NA	NA	NA
T1M6	1	both	C	Sing	KRP	NA	envmt	Soil	soil	NA	NA	NA
T1M7	1	both	C	Sing	KRP	NA	envmt	Soil	soil	NA	NA	NA
T1M8	1	both	C	Sing	KRP	NA	envmt	Soil	soil	NA	NA	NA
T2M1	1	both	C	Sing	BTNP	NA	envmt	Soil	soil	NA	NA	NA
T2M2	1	both	C	Sing	BTNP	NA	envmt	Soil	soil	NA	NA	NA
T2M3	1	both	C	Sing	BTNP	NA	envmt	Soil	soil	NA	NA	NA
T2M4	1	both	C	Sing	BTNP	NA	envmt	Soil	soil	NA	NA	NA
T2M5	1	18S	C	Sing	BTNP	NA	envmt	Soil	soil	NA	NA	NA
T2M6	1	both	C	Sing	BTNP	NA	envmt	Soil	soil	NA	NA	NA
T2M7	1	both	C	Sing	BTNP	NA	envmt	Soil	soil	NA	NA	NA
T2M8	1	both	C	Sing	BTNP	NA	envmt	Soil	soil	NA	NA	NA
T3M1	1	both	C	Sing	UPR	NA	envmt	Soil	soil	NA	NA	NA
T3M2	1	both	C	Sing	UPR	NA	envmt	Soil	soil	NA	NA	NA
T3M3	1	both	C	Sing	UPR	NA	envmt	Soil	soil	NA	NA	NA
T3M4	1	both	C	Sing	UPR	NA	envmt	Soil	soil	NA	NA	NA
T3M5	1	both	C	Sing	UPR	NA	envmt	Soil	soil	NA	NA	NA
T3M6	1	both	C	Sing	UPR	NA	envmt	Soil	soil	NA	NA	NA
T3M7	1	both	C	Sing	UPR	NA	envmt	Soil	soil	NA	NA	NA
T3M8	1	both	C	Sing	UPR	NA	envmt	Soil	soil	NA	NA	NA
UNa02	2	both	C	Sing	BTNP	NA	tube	Tube	tube	squat	5.5	1.6
UNa04	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	5.5	3.4
UNa05	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	5.75	4.6
UNa06	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	5.75	1.5
UNg01	2	both	C	Sing	BTNP	NA	tube	Tube	tube	squat	5.5	7.4
UNg02	2	both	C	Sing	BTNP	NA	tube	Tube	tube	squat	7.5	5.1
UNg04	2	16S	C	Sing	UPR	NA	tube	Tube	tube	squat	5.5	9.6
UNg07	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	5	0.4
UNg08	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	7.5	5.4
UNg09	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	5.75	0.5
UNr01	2	both	C	Sing	BTNP	NA	tube	Tube	tube	squat	5.5	2.6
UNr02	2	both	C	Sing	BTNP	NA	tube	Tube	tube	squat	5.5	10.5
UNr03	2	both	C	Sing	BTNP	NA	tube	Tube	tube	squat	5.5	6.2
UNr04	2	both	C	Sing	BTNP	NA	tube	Tube	tube	squat	7	7

Supplementary Table 3.1 (Continued)

SampleID	Seq Set	16S/ 18S	Pro ject	Loc.	Site	Exp Day	Type_ 1	Type _2	Type_3	Form	pH	Vol.
UNr05	2	16S	C	Sing	BTNP	NA	tube	Tube	tube	squat	7.5	0.9
UNr10	2	both	C	Sing	UPR	NA	tube	Tube	tube	squat	5	5
UNr11	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	5.5	0.5
UNr12	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	5.5	1
UNr13	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	5.5	0.5
UNr14	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	5	3.1
UNr15	2	both	C	Sing	KRP	NA	tube	Tube	tube	squat	5.25	3.9
UPRLwfilt	1	both	C	Sing	UPR	NA	envmt	Leaf	leaf	NA	NA	NA

Legend: C = Comparison, E = Experiment, GC = Gulf Coast, HF = Harvard Forest, MB = Maliau Basin, Sing = Singapore, TS = Tom Swamp.

Supplementary Table 3.2: Shannon Diversity (mean and standard deviations)

	Bacteria	Eukaryota
Soil	9.3 (+/- 0.7)	5.5 (+/- 1.4)
Bog	8.3 (+/- 0.9)	5.2 (+/- 1.2)
Nepenthes	5.2 (+/- 1.8)	3.1 (+/- 0.9)
Sarracenia	4.6 (+/- 1.2)	2.2 (+/- 0.9)

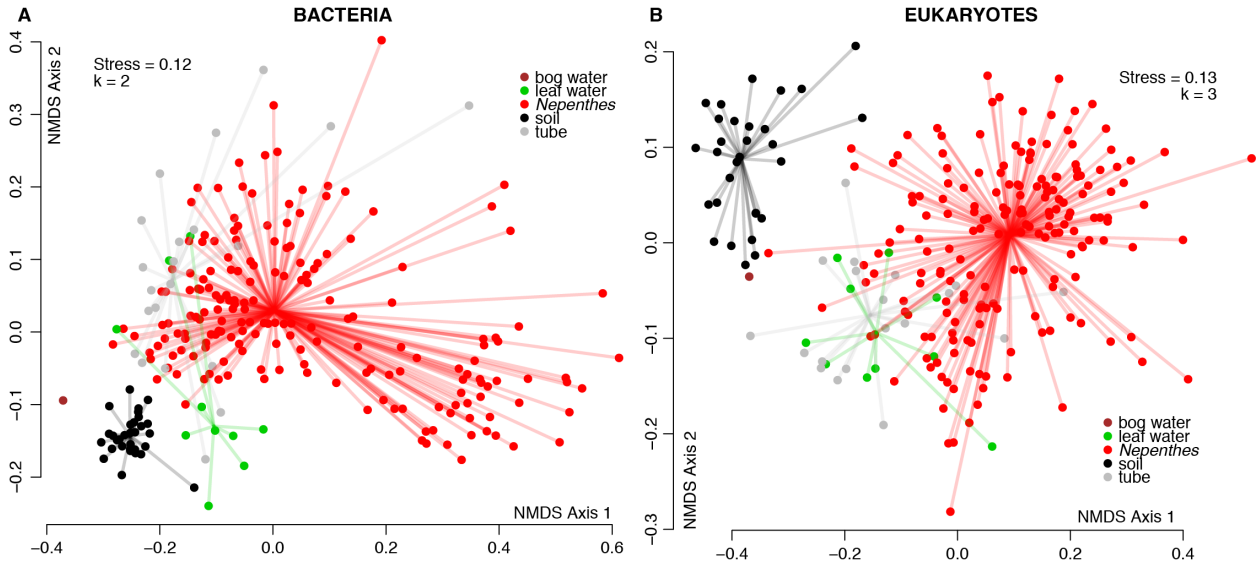
Supplementary Table 3.3: Comparing the Beta-diversity of communities

		Bacterial unweighted UniFrac				Eukaryotic unweighted UniFrac			
		adonis / mantel		envfit / ordisurf		adonis / mantel		envfit / ordisurf	
Group	Factor	R^2 / r	P	R^2	P	R^2 / r	P	R^2	P
All samples (from Figure 1C)	Pitchers vs. environmental samples	0.08	< 0.001	0.31	< 0.001	0.08	< 0.001	0.38	< 0.001
	Region (HF, GC, S, and MB)	0.09	< 0.001	0.21	< 0.001	0.13	< 0.001	0.29	< 0.001
Nepenthes samples	pH (range: 1.5 - 7.5)	0.63	< 0.001	0.74	< 0.001	0.17	< 0.001	0.21	< 0.001
	Volume (range: 0.01 - 500 mL)	0.04	0.207	0.07	0.015	-0.04	0.803	0.09	0.004
	Pitcher species (8 species)	0.18	< 0.001	0.38	< 0.001	0.15	< 0.001	0.24	< 0.001
	Collection site (4 sites)*	0.03	< 0.001	0.09	< 0.001	0.04	< 0.001	0.13	< 0.001
Sarracenia samples	pH (range: 3 - 8.5)	0.11	0.006	0.13	< 0.001	-0.05	0.852	0.01	0.152
	Volume (range: 0.1 - 43 mL)	0.15	0.007	0.31	< 0.001	0.17	0.002	0.17	< 0.001
	Pitcher species (6 species)	0.17	< 0.001	0.39	< 0.001	0.21	< 0.001	0.42	< 0.001
	Pitcher growth form (squat vs tall)	0.08	< 0.001	0.31	< 0.001	0.21	< 0.001	0.39	< 0.001
	Collection site (14 sites)*	0.12	< 0.001	0.33	< 0.001	0.09	< 0.001	0.32	< 0.001
Natural *Nepenthes* and *Sarracenia* samples	Genus	0.09	< 0.001	0.33	< 0.001	0.14	< 0.001	0.56	< 0.001
Natural *Nepenthes* with experimental samples	pH	0.05	< 0.001	0.67	< 0.001	0.12	< 0.001	0.23	< 0.001
	Region (HF, S, and MB)	0.07	< 0.001	0.19	< 0.001	0.12	< 0.001	0.41	< 0.001

Adonis (Permutation MANOVAs) and mantel tests are done on the dissimilarity matrices, while envfit and ordisurf are done on the NMDS results. Adonis and envfit are for categorical variables, while mantel and ordisurf are for continuous variables.
Legend: HF = Harvard Forest, GC = Gulf Coast, S = Singapore, MB = Maliau Basin.
*In these calculations, the adonis test controls for differences in pitcher species among the sites.

Supplementary Table 3.4: Comparison metagenomes

Description	Type	Source	ID
Damari_Lake	lake	SRA	ERR358545
Erken_Lake	lake	SRA	ERR358542
Mendota_Lake	lake	SRA	ERR358549
Trout_Lake	lake	SRA	ERR358547
A_thaliana_phyllo	phyllosphere	MG-RAST	4447810.3
Clover_phyllo	phyllosphere	MG-RAST	4447811.3
Corn_phyllo	phyllosphere	SRA	SRR2924445
Rice_phyllo	phyllosphere	MG-RAST	4450328.3
Soybean_phyllo	phyllosphere	MG-RAST	4441205.3
Forest_soil	soil	MG-RAST	4477899.3
Grassland_soil1	soil	MG-RAST	4511045.3
Grassland_soil2	soil	MG-RAST	4477804.3
Mangrove_rhizo	soil	MG-RAST	4535147.3
Mangrove_soil	soil	MG-RAST	4506447.3
Meadow_soil	soil	MG-RAST	4449357.3
Rotham_soil	soil	MG-RAST	4453436.3
Tropical_soil1	soil	MG-RAST	4477807.3
Tropical_soil2	soil	MG-RAST	4477875.3
Tundra_soil	soil	MG-RAST	4477874.3



Supplementary Figure 3.1: NMDS plots of bacterial (A) and eukaryotic (B) communities in different habitats in Southeast Asia.