Austral ECOLOGY A Journal of ecology in the Southern Hemisphere

Austral Ecology (2015) ••, ••–••

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

LEONORA S. BITTLESTON,[1,*] CHRISTOPHER C. M. BAKER,[1] LILA B. STROMINGER,[1] ANNE PRINGLE[2] AND NAOMI E. PIERCE[1]

[1]*Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138 (Email: lbittles@fas.harvard.edu), and* [2]*Harvard Forest, Harvard University, Petersham, Massachusetts, USA*

Abstract The biodiversity of tropical forests consists primarily of small organisms that are difficult to detect and characterize. 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 an 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 characterize 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 specialization of certain arthropod fauna.

Key words: 18S amplicon sequencing, carnivorous plant, insect, 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 characterize their communities using traditional survey methods. Recent efforts to describe the full diversity of arthropods in a Panamanian tropical rainforest found 6144 species in less than one-half hectare, and estimated that 25 000 arthropod species exist within a 6000-ha reserve (Basset *et al.* 2012). The rainforests of Southeast Asia are currently threatened by anthropogenic activities, including the highest relative rates of deforestation compared with 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 1800s (Riley 1874) and are now model systems for food web and microcosm studies (Kitching 2000; Kneitel & 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 & 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,

*Corresponding author.
Accepted for publication April 2015.

doi:10.1111/aec.12271

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 & 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, although 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' distribution (Beaver 1985). The inquilines vary with host species (Clarke & Kitching 1993) and have complex predator–prey dynamics (Mogi & Yong 1992). Some species of dipteran insects and aquatic mites appear to be specialized to *Nepenthes* habitats (Ratsirarson & Silander 1996; Fashing 2002; Fashing & 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 characterize 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 characterizing community structure. Sequences are typically shorter than Sanger-sequenced barcodes, and thus contain less taxonomic information. Polymerase Chain Reaction (PCR) biases can affect final sequence abundance, 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. 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 with rRNA genes (Deagle *et al.* 2014).

To the best of our knowledge, this study is the first attempt to characterize 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 characterization of eukaryotic communities, and specifically arthropods? Second, how does 18S metabarcoding compare with 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 specialization 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 1). Pitcher fluid was collected from the three species of *Nepenthes* with natural distributions in Singapore: *N. gracilis*, *N. rafflesiana* and *N. ampullaria* (Fig. 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 Appendix S1). 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 cetyltrimethylammoniumbromide 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.

doi:10.1111/aec.12271

Table 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

BTNP, Bukit Timah Nature Preserve; KRP, Kent Ridge Park; UPR, Upper Peirce Reservoir Park.

DNA extraction and 18S amplicon NGS

To concentrate cells, half of the fluid from each sample was either filtered through sterilized 0.22-μm 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 were done according to the Earth Microbiome Project protocols (http://www.earthmicrobiome.org/emp-standard-protocols/18s).

18S quality control and operational taxonomic unit (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 1, Appendix S1). DNA sequences, averaging 151 nucleotides in length, were clustered into OTUs (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 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 (Fig. 1a), a rarefaction curve (Fig. 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: 1595. We then took the square root of all observations to decrease the impact of certain OTUs having falsely high abundance 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 programme 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; however, many of our sequences had no matches, or matched to unnamed sequences. Consequently, sequences were assigned taxonomy using BLASTn and the National Center for Biotechnology Information (NCBI) database (Appendix S2), and aligned with the programme 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) (Fig. 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

 doi:10.1111/aec.12271

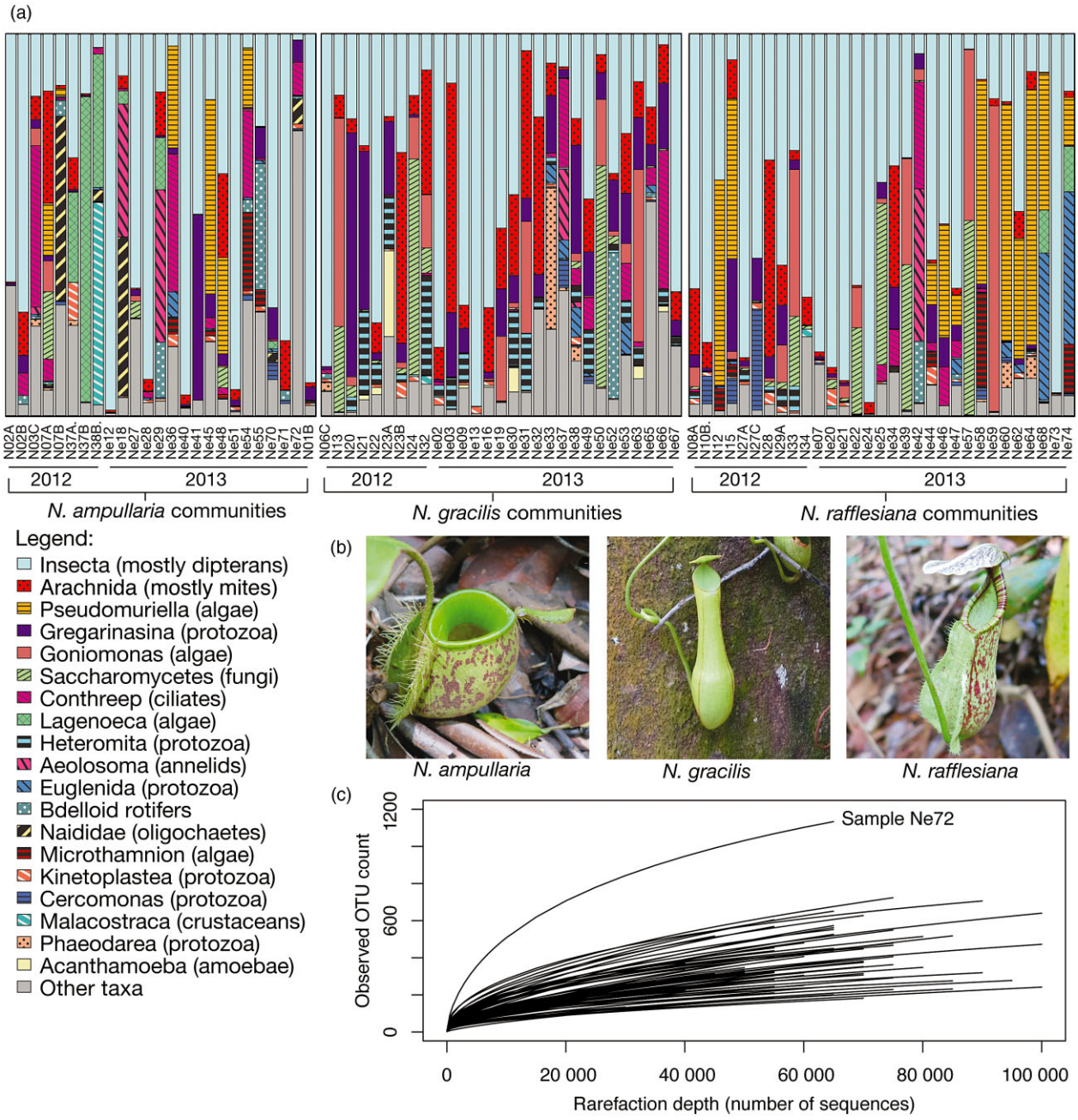

Fig. 1. Eukaryotes in *Nepenthes* pitcher plants. (a) Summary of the eukaryotic taxa found within the pitcher fluid of three *Nepenthes* species. Each column is one sample, and the y-axis is the relative sequence abundance of the taxa listed in the legend. Samples are grouped by *Nepenthes* species, shown in the photographs (b). (c) Rarefaction curves for each sample showing the number of observed operational taxonomic unit (OTU) (y-axis) at different sampling depths (x-axis). Sample Ne72 contained soil and was removed from subsequent analyses.

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 (Appendix S2). Sequences were aligned and a tree was built as described above for the COI sequences (Fig. 2).

Arthropod counts and comparison with 18S sequences

Arthropods from a subset of the 2012 samples (Table 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*

doi:10.1111/aec.12271

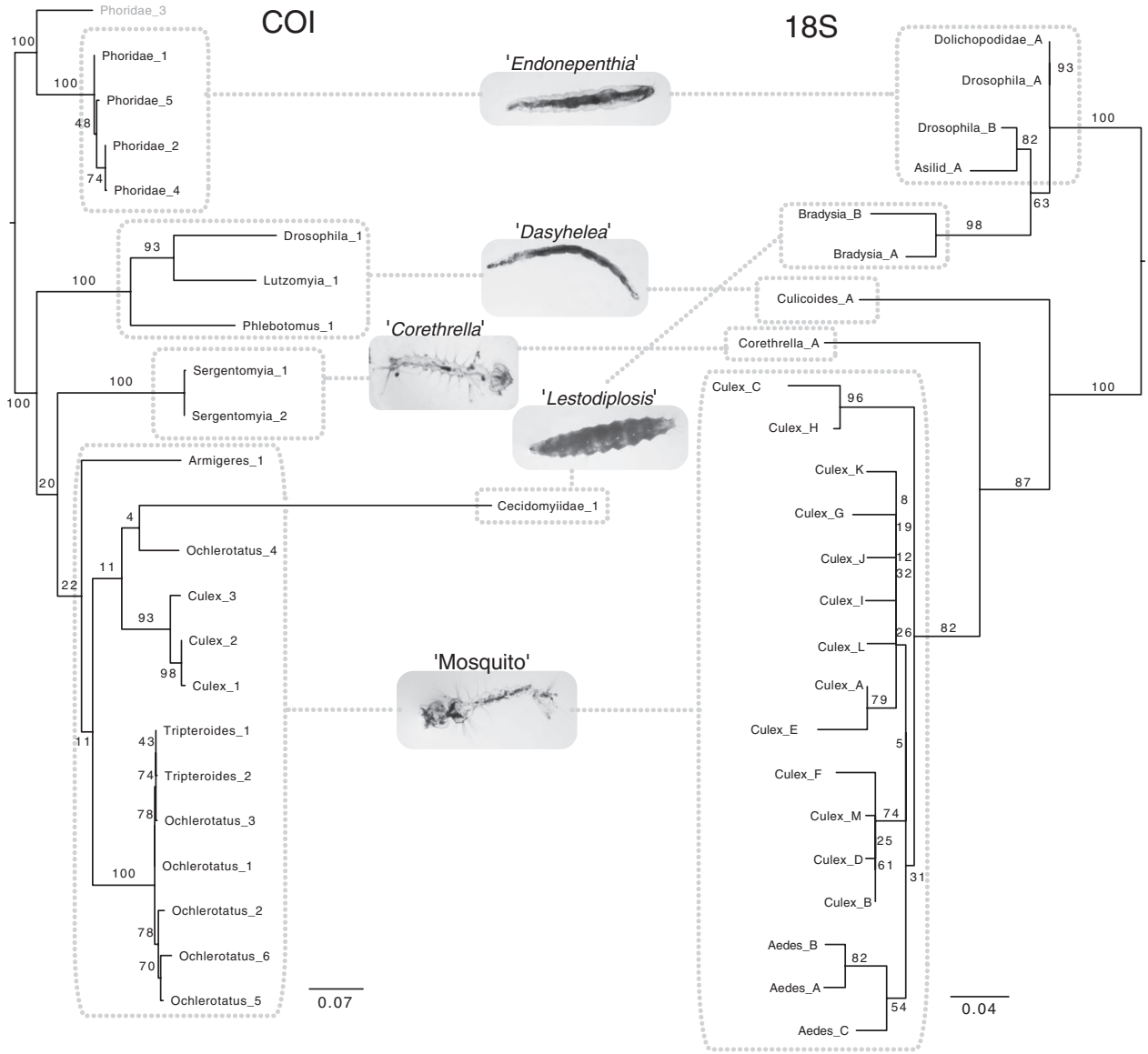

Fig. 2. COI and 18S maximum likelihood phylogenetic trees with bootstrap values at each node. The clades corresponding to each dipteran morphospecies are outlined and connected across the phylogenies. Taxa on the branch tips are named according to best BLASTn matches.

(Kai Lok *et al.* 1997). We also took photographs of inquilines under a dissecting microscope for future reference.

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 with the number of 18S sequences in the OTUs in each sample (Table 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 (Fig. 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 (Fig. 3b).

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*

 doi:10.1111/aec.12271

Table 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 operational taxonomic unit table.

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 specialization 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 three 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 specialization level with very low-abundance observations. The network was built and graphed using the bipartite package in R (Dormann *et al.* 2009) (Fig. 4). Network-level specialization 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 specialization) to 1 (all organisms completely specialized). We used a null model based on the Patefield algorithm to test if the network is more specialized 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 three 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 Appendix S3 and associated methods).

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 7229 OTUs (3115 were

doi:10.1111/aec.12271

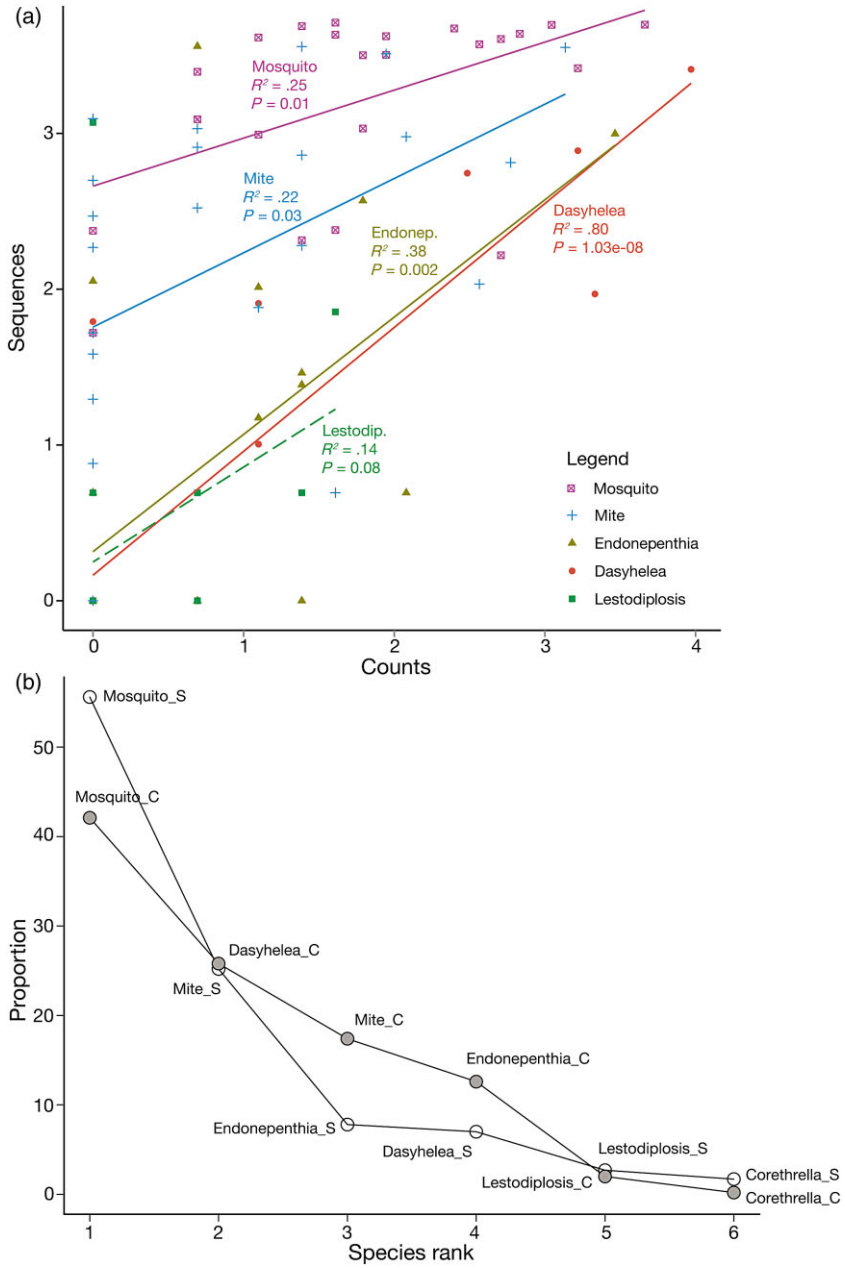

Fig. 3. Comparing individuals counts with 18S sequences. (a) Scatter plot of inquiline individual counts and 18S sequences plotted on a $\log_{10}$-$\log_{10}$ scale. Regression lines and *P*-values from the permutational linear models overlie the scatter plot. Solid lines are significant. (b) Rank abundance 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 labelled with a 'C' and grey circles.

singletons). Per sample sequence counts for arthropods ranged from 1595 to 65 604.

Rarefaction curves for each sample of all observed eukaryotic OTUs plotted against sequences per sample appear to be levelling off (Fig. 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 (Fig. 1c), with many OTUs observed only once. Ne72 appeared to have soil inside when collected. It possesses a much higher observed species richness (Fig. 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 (Fig. 1a). The most common arthropods in samples are dipteran

 doi:10.1111/aec.12271

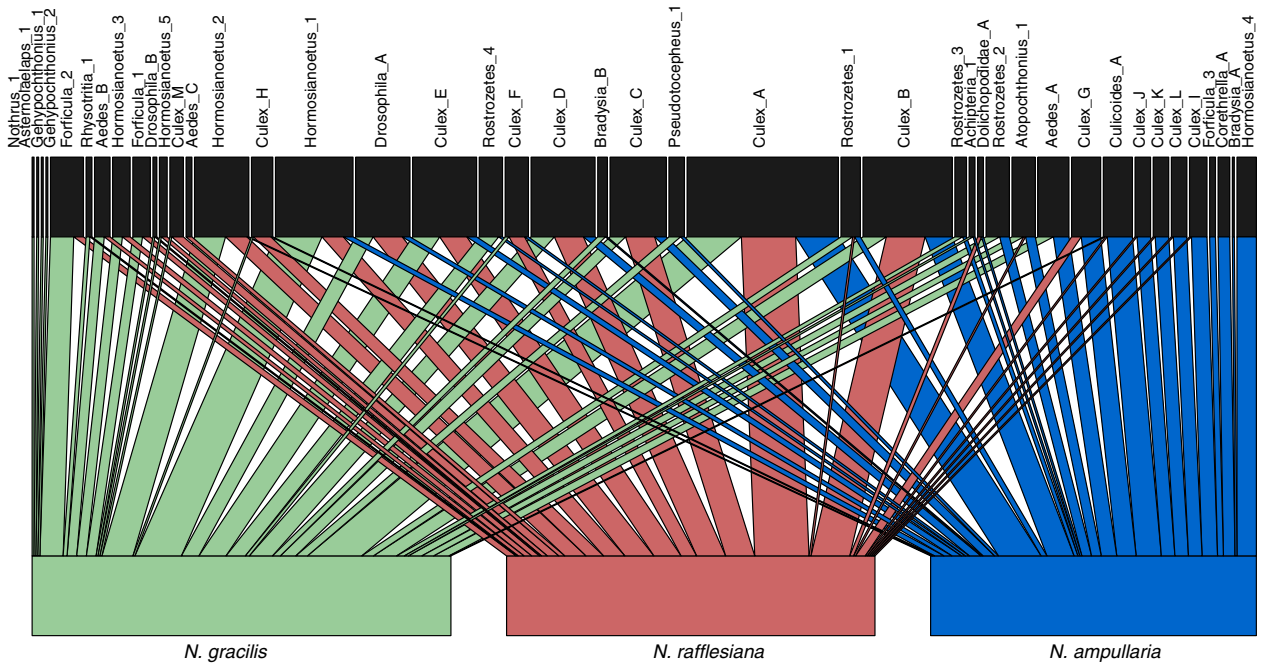

Fig. 4. Bipartite network of inquilines and *Nepenthes* host species. Inquiline operational taxonomic units (OTUs) are listed above, dipteran names are followed by letters, and mites names are followed by numbers. The graph is organized to have the fewest crossing lines; OTUs associated with only one host species are towards the edges, while shared OTUs are in the centre. Observations of three 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.

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 (Fig. 1a, Appendix S4). 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 (Fig. 1a, Appendix S4). 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, 23 unique COI barcodes of dipteran inquilines were assigned taxonomy (Appendix S2) and visualized in a phylogenetic tree (Fig. 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 (Appendix S2). 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* (Fig. 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

doi:10.1111/aec.12271

Table 3. Taxonomy assignments and correspondence between COI and 18S

Morphospecies name		COI taxonomy assignment		18S taxonomy assignment		
Genus	Family	Genus	Family	Genus	Family	Correspondence
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

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% (Appendix S2).

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* (Fig. 2).

The *Corethrella* morphospecies was represented by one OTU, with taxonomic assignment to the genus *Corethrella* at 100% sequence similarity (Appendix S2). In the 18S tree, the *Corethrella* OTU falls sister to the mosquito clade (Fig. 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 (Fig. 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 (Appendix S2).

Comparing COI and 18S

The COI and 18S trees correspond well, with the same placement for four of the five dipteran insect morphospecies (Fig. 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 is therefore 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 (Fig. 2).

Taxonomic assignments of COI and 18S sequences corresponded with morphospecies taxonomy for some insects but not for others (Table 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 3).

Arthropod counts and comparisons with 18S sequences

Individual counts of inquilines correlate with 18S sequence counts, although the variance explained by regressions is low (Fig. 3a). Most correlations are significant. The strongest correlation is for the *Dasyhelea*

doi:10.1111/aec.12271

morphospecies: $R^2 = 0.80$, $P = 1.03e\text{-}08$. 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 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 (Fig. 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 beta dispersion of host species groups was 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 (Fig. 4) illustrates how certain OTUs are present in samples from only one host species (e.g. 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 the 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 (Fig. 4). The network has a relatively low level of specialization: $H_2' = 0.26$, but it is highly significant when compared with a null model: $P < 0.001$. Our results are robust, as we found all networks to be significantly specialized, 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 specialization as the network of inquilines alone: $H_2' = 0.25$, $P < 0.001$.

A spring-embedded network (Appendix S3) 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 (Fig. 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 characterization 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 characterizing eukaryotic communities within pitcher plants, especially communities with arthropods. Insects and mites were the most highly represented organisms in our samples (Fig. 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 (Fig. 1a, Appendix S4), 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, 2015). NGS provides a molecular window through which to observe buried or microscopic organisms.

Perhaps surprisingly, the majority of our sequences were from insect DNA (Fig. 1a, Appendix S4), but not ants. Instead, most sequences were from the inquiline insects living within *Nepenthes* pitchers. DNA may be

doi:10.1111/aec.12271

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, the 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 with 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 '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 3). Per cent identity of both 18S and COI sequences compared with the best BLASTn hit is often very low, with the lowest being 83% for 18S and 84% for COI (Appendix S2). 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 (Fig. 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 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 abundance does not correspond for inquilines of intermediate abundance (Fig. 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 abundance 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). 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.

 doi:10.1111/aec.12271

Can 18S metabarcoding be used to study ecological dynamics and host specialization 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 specialization of certain inquilines to particular hosts, as well as the presence of core inhabitants in all three *Nepenthes* species (Fig. 4). The bipartite network of arthropods in *Nepenthes* pitcher microcosms is significantly specialized across the three host species, according to null models of the H_2' network-level specialization 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 2 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 (Appendix S3). 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, 2015). This will further illuminate degrees of specialization 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.

ACKNOWLEDGEMENTS

Our research was supported by an NSF Graduate Research Predoctoral Fellowship and NSF Doctoral Dissertation Improvement Grant DEB-1400982 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.

REFERENCES

Acinas S. G., Sarma-Rupavtarm R., Klepac-Ceraj V. & Polz M. F. (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* **71,** 8966–9.

Adlassnig W., Peroutka M. & Lendl T. (2011) Traps of carnivorous pitcher plants as a habitat: composition of the fluid, biodiversity and mutualistic activities. *Ann. Bot.* **107,** 181–94.

Albert V. A., Williams S. E. & Chase M. W. (1992) Carnivorous plants: phylogeny and structural evolution. *Science* **257,** 1491–5.

Altschul S. F., Gish W., Miller W., Myers E. W. & Lipman D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–10.

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.

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

Anderson M. J. (2006) Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62,** 245–53.

Basset Y., Cizek L., Cuénoud P. *et al.* (2012) Arthropod diversity in a tropical forest. *Science* **338,** 1481–4.

Beaver R. A. (1985) Geographical variation in food web structure in *Nepenthes* pitcher plants. *Ecol. Entomol.* **10,** 241–8.

Bik H. M., Porazinska D. L., Creer S., Caporaso J. G., Knight R. & Thomas W. K. (2012) Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol. Evol. (Amst.)* **27,** 233–43.

Blüthgen N., Menzel F. & Blüthgen N. (2006) Measuring specialization in species interaction networks. *BMC Ecol.* **6,** 9. doi: 10.1186/1472-6785-6-9

Caporaso J. G., Kuczynski J., Stombaugh J. *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Meth.* **7,** 335–6.

Caporaso J. G., Lauber C. L., Walters W. A. *et al.* (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* **108** (Suppl 1), 4516–22.

Caporaso J. G., Lauber C. L., Walters W. A. *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6,** 1621–4.

Chapin C. T. & Pastor J. (1995) Nutrient limitations in the northern pitcher plant *Sarracenia purpurea*. *Can. J. Bot.* **73,** 728–34.

Clarke C. M. & Kitching R. L. (1993) The metazoan food webs from six Bornean *Nepenthes* species. *Ecol. Entomol.* **18,** 7–16.
Clarke K. R. (1993) Non-parametric multivariate analyses of changes in community structure. *Austral Ecol.* **18,** 117–43.
Deagle B. E., Jarman S. N., Coissac E., Pompanon F. & Taberlet P. (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol. Lett.* **10,** 20140562.
Dormann C. F. (2011) How to be a specialist? Quantifying specialisation in pollination networks. *Network Biol.* **1,** 1–20.
Dormann C. F., Fründ J., Blüthgen N. & Gruber B. (2009) Indices, graphs and null models: analyzing bipartite ecological networks. *Open Ecol. J.* **2,** 7–24.
Edgar R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–7.
Edgar R. C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–1.
Ellison A. M. & Gotelli N. J. (2009) Energetics and the evolution of carnivorous plants – Darwin's 'most wonderful plants in the world'. *J. Exp. Bot.* **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. *Aust. J. Entomol.* **41,** 7–17.
Fashing N. J. & Chua T. H. (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. *Int. J. Acarol.* **28,** 157–67.
Folmer O., Black M., Hoeh W., Lutz R. & Vrijenhoek R. (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Mol. Marine Biol. Biotechnol.* **3,** 294–9.
Gronemeyer T., Coritico F., Wistuba A. *et al*. (2014) Four new species of *Nepenthes* L. (Nepenthaceae) from the central mountains of Mindanao, Philippines. *Plants* **3,** 284–303.
Hajibabaei M., Shokralla S., Zhou X., Singer G. A. C. & Baird D. J. (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* **6,** e17497.
Hatcher M. J., Dick J. T. A. & Dunn A. M. (2012) Diverse effects of parasites in ecosystems: linking interdependent processes. *Front. Ecol. Environ.* **10,** 186–94.
Hebert P. D. N., Cywinska A., Ball S. L. & deWaard J. R. (2003) Biological identifications through DNA barcodes. *Proc. Biol. Sci.* **270,** 313–21.
Huse S. M., Welch D. M., Morrison H. G. & Sogin M. L. (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* **12,** 1889–98.
Kai Lok C., Chen L. M. J., Choo J. P. S. *et al*. (1997) *A Guide to the Carnivorous Plants of Singapore* (ed. H. T. W. Tan). Singapore Science Centre, Singapore.
Kitching R. L. (2000) *Food Webs and Container Habitats*. Cambridge University Press, Cambridge.
Klappenbach J. A., Saxman P. R., Cole J. R. & Schmidt T. M. (2001) rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res.* **29,** 181–4.
Kneitel J. M. & Miller T. E. (2002) Resource and top-predator regulation in the pitcher plant (*Sarracenia purpurea*) inquiline community. *Ecology* **83,** 680–8.
Mogi M. & Yong H. S. (1992) Aquatic arthropod communities in *Nepenthes* pitchers: the role of niche differentiation, aggregation, predation and competition in community organization. *Oecologia* **90,** 172–84.
Moran J. A., Booth W. E. & Charles J. K. (1999) Aspects of pitcher morphology and spectral characteristics of six Bornean *Nepenthes* pitcher plant species: implications for prey capture. *Ann. Bot.* **83,** 521–8.
Oksanen J., Blanchet F. G., Kindt R. *et al*. (2013) vegan: community ecology package (online). [Cited 9 March 2015.] Available from URL: http://CRAN.R-project.org/package=vegan
Prokopowich C. D., Gregory T. R. & Crease T. J. (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46,** 48–50.
Pruesse E., Quast C., Knittel K. *et al*. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–96.
R Core Team (2014) R: a language and environment for statistical computing (online). [Cited 9 March 2015.] Available from URL: http://www.R-project.org/
Ratsirarson J. & Silander J. A. (1996) Structure and dynamics in *Nepenthes madagascariensis* pitcher plant micro-communities. *Biotropica* **28,** 218–27.
Riley C. V. (1874) On the insects more particularly associated with *Sarracenia variolaris* (spotted trumpet-leaf). *Can. Entomol.* **6,** 207–14.
Shokralla S., Singer G. A. C. & Hajibabaei M. (2010) Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. *Biotechniques* **48,** 233–4.
Sodhi N. S., Koh L. P., Brook B. W. & Ng P. K. L. (2004) Southeast Asian biodiversity: an impending disaster. *Trends Ecol. Evol. (Amst.)* **19,** 654–60.
Srivastava D. S., Kolasa J., Bengtsson J. *et al*. (2004) Are natural microcosms useful model systems for ecology? *Trends Ecol. Evol. (Amst.)* **19,** 379–84.
Stamatakis A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–13.
Stoeck T., Bass D., Nebel M. *et al*. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19** (Suppl 1), 21–31.
Taberlet P., Coissac E., Pompanon F., Brochmann C. & Willerslev E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21,** 2045–50.
Yu D. W., Ji Y., Emerson B. C. *et al*. (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol. Evol.* **3,** 613–23.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. *Nepenthes* samples and metadata.
Appendix S2. The most abundant eukaryotic taxa from the *Nepenthes* samples.
Appendix S3. BLASTn taxonomy assignments of 18S and COI sequences.
Appendix S4. Spring-embedded network of arthropod OTUs from *Nepenthes* pitchers.

 doi:10.1111/aec.12271