Cascading effects of habitat loss on ectoparasite communities and their associated bacterial microbiomes

Kelly A. Speer[1], Tiago Teixeira[2], Alexis Brown[3], Susan Perkins[1], Katharina Dittmar[4], Melissa Ingala[1], Claudia Wultsch[1], Konstantinos Krampis[5], Carl Dick[6], Spencer Galen[7], Nancy Simmons[1], and Elizabeth Clare[2]

[1]American Museum of Natural History
[2]Queen Mary University of London
[3]Stony Brook University Department of Ecology and Evolution
[4]State University of New York at Buffalo
[5]Hunter College CUNY
[6]Western Kentucky University
[7]The Academy of Natural Sciences of Drexel University

July 5, 2022

Abstract

Suitable habitat fragment size, isolation, and distance from a source are important variables influencing community composition of plants and animals, but the role of these environmental factors in determining composition and variation of host-associated microbial communities is poorly known. In parasite-associated microbial communities, it is hypothesized that evolution and ecology of an arthropod parasite will influence its microbiome more than broader environmental factors, but this hypothesis has not been extensively tested. To examine the influence of the broader environment on the parasite microbiome, we applied high-throughput sequencing of the V4 region of 16S rRNA to characterize the microbiome of 222 obligate ectoparasitic bat flies (Streblidae and Nycteribiidae) collected from 155 bats (representing six species) from ten habitat fragments in the Atlantic Forest of Brazil. Parasite species identity is the strongest driver of microbiome composition. To a lesser extent, reduction in habitat fragment area, but not isolation, is associated with an increase in connectance and betweenness centrality of bacterial association networks driven by changes in the diversity of the parasite community. Controlling for the parasite community, bacterial network topology covaries with habitat patch area and exhibits parasite-species specific responses to environmental change. Taken together, habitat loss may have cascading consequences for communities of interacting macro- and microorgansims.

Introduction

Deforestation has well-documented, devastating consequences on species survival [1], global warming [2], and zoonotic disease emergence [3, 4]. This is particularly worrying in tropical forests, which lost 6% of in their global area between 1990 and 2015 [5]. For example, the Atlantic Forest of Brazil, one of the world's most biodiverse regions, occupies only 28% of its original extent [6]. The consequences of this habitat loss have primarily been examined in macroorganisms, but we do not yet understand the extent to which microorganisms like bacteria, viruses, fungi, and single-cell eukaryotes respond to deforestation, especially those microorganisms that are obligately associated with a living host. These microorganisms are integral members of wildlife communities; changes in their presence or abundance may alter the function of communities [7], the health of individuals and species [8], and the transmission of pathogens between members of the community [9].

Here, we examine the bacterial microbiome within ectoparasites of bats in forest fragments of varying area and isolation as a model for testing the hypothesis that habitat loss affects host-associated microbial communities (Figure 1A). We use island biogeography theory as a null hypothesis for the way we expect microbiomes to behave if the environment is driving community composition instead of the host. This theory states that small, isolated habitats will support low-diversity communities that are a subset of the species found in larger source communities [10]. Historically, studies of the island biogeography of parasites have treated hosts as suitable habitat, not the broader environment where the host and parasite live [11, 12]. There are two limitations to this thinking: only an extremely small subset of parasites spend their entire lifespan on the body of a single host individual, and this idea ignores the impact that the environment has on parasite microbiomes that may subsequently impact parasite survival [13]. Hosts are not islands that perfectly constrain their parasites, and parasites are not islands that perfectly constrain their microbes. Previous research suggests that the environment influences microbiome composition following expectations of island biogeography [14, 15]. In other cases, the environment does not dictate microbiome composition [16, 17]. Variation in the ability of the environment to filter members of the microbiome community is a reflection of the complexity and diversity of microorganisms themselves [16, 18]. This diversity makes it difficult to tease apart the impacts of host and environment on microbiome community composition.

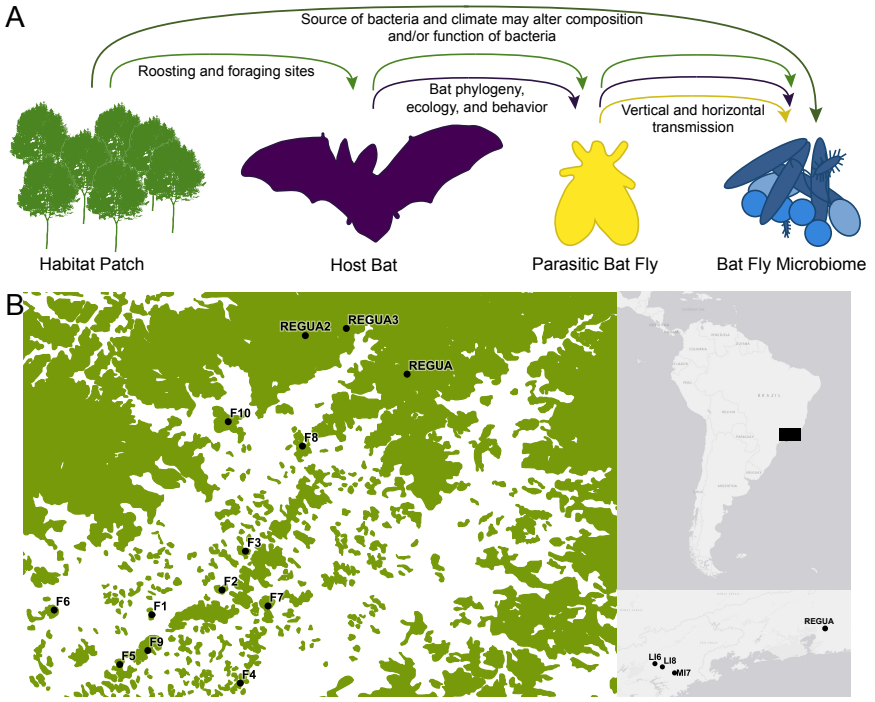


Figure 1: Sampling Design– Illustration of how environment ("habitat patch"), host bat, and parasitic bat fly each influence the microbiomes within bat flies (A; see in-text description of line colors). Sampling map constructed in QGIS v3.12 of REGUA area sites with fragments outside of REGUA labelled with the prefix "F" and ordered from smallest (F1) to largest (F10; B), the extent of the sampled area is shaded in black in the map at the top right, and the southern sites are mapped in relation to REGUA in the map on the bottom right. Green area of the map on the left indicates forested habitat based on imagery from SOS Mata Atlántica, while white areas are all non-forested habitat types.

Obligate parasites represent a convenient system within which to test potential drivers of microbiome variation for several reasons. The number of factors influencing variation may be more limited in the microbiomes

of obligate parasites than in environmental microbiomes (e.g., soil or water) or microbiomes of free-living host species. Unlike free-living host species, obligate ectoparasites have extremely specialized diets, and their movements in the broader environment beyond their host are constrained by their dependence on a host to survive. Ectoparasitic arthropods also have characteristically depauperate microbiome communities compared to arthropods with diverse diets [19]. We can leverage the hierarchical nature of the host-parasite-microbiome system to clearly delimit the microbiome community and restrict the sources of colonizing bacteria that may invade the parasite microbiome, hence providing a manageable system for testing hypotheses about community composition and factors governing assembly of the microbiome.

In this study, we used bat flies (Diptera: Streblidae and Nycteribiidae), which are obligate blood-feeding ectoparasites of bats, to assess community composition of parasite-associated microorganisms across a fragmented landscape in the Atlantic Forest. The microbiome of bat flies may be influenced by the parasitic bat fly, the host bat, and the environment in several ways (Figure 1A). First, bat fly microbiomes may be vertically inherited or horizontally acquired, as in other insect-microbiome associations (Figure 1A, yellow line; [20]. Second, bat flies are host-specific and depend on their host for dispersal [21, 22], meaning that the host bat may also influence the bat fly microbiome by altering the community of bat flies from which bacteria may be horizontally acquired (Figure 1A, dark purple line). For example, bat maternity colonies likely support more abundant parasite communities than bachelor colonies, because bat flies preferentially parasitize female and young bats [23]. By supporting a smaller community of bat flies, bachelor colonies may decrease the sources from which bat flies horizontally acquire bacteria, leading to host-sex-based variation in the microbiome. Other aspects of bat phylogeny, ecology, and behavior may similarly influence bat fly microbiomes, including roost preference, feeding guild, and host bat species identity. Third, habitat patches support a specific bat community based on the availability of roosting and foraging sites, which subsequently alters the diversity and abundance of bat flies [24]. These changes in the local bat fly community may be reflected in associated microbiomes. Lastly, the environment may directly impact bat fly microbiome composition (Figure 1A, dark green line). Bat flies and all other members of the Hippoboscoidea are adenotrophically viviparous, a condition in which a single egg hatches inside the female fly and the larva feeds from milk glands until it is ready to pupate [21]. In the case of bat flies, the female fly leaves the host bat to deposit the larva on the roost substrate [21], providing opportunities for the environment to act as a source of bacteria for the microbiome of bat flies (Figure 1A). Beyond acting as a source of bacteria for the microbiome, deforestation may increase the local temperature of small habitat patches and directly impact arthropod-associated microbiomes due to thermal constraints of some bacteria [25, 26]. Using the mosaic landscape of the Atlantic Forest, we can examine whether bat fly-associated microbiomes respond to environmental change following island biogeography theory or whether the host bat and parasitic bat fly more strongly determine bat fly microbiome composition.

Methods

Sample collection and landscape metrics

Bat flies were collected from bats in 11 habitat patches of the Atlantic forest of Brazil, State of Rio de Janeiro from 18 December 2015 to 19 January 2017 (Table 1, Figure 1B, Table S1, Figure S1; [27], including a large protected area of pristine and secondary forest belonging to the Reserva Ecológica de Guapiaçu (REGUA). REGUA is the third largest remaining expanse of Atlantic Forest and was sampled in three separate locations. Samples were additionally collected from three geographically distant habitat fragments (southern sites; Figure 1B). Each site was sampled for 6 nights, 6 hours per night or at least 2 hours if there was heavy rain, and between 7 and 10 ground-level mist nets were used to capture bats each night (approximately 60m of nets were set per night; [27]. Bats were removed from mist nets and placed into freshly washed cloth bags for holding and to minimize cross-contamination of ectoparasites. Each bat was searched for approximately 45s for ectoparasites, which were captured using featherweight forceps and immediately transferred to tubes containing 92% ethanol, stored at room temperature overnight, and subsequently transferred to -20°C. Bats were identified in the field following [28, 29]. All capture and handling methods followed recommendations

in [30]. Because many bat species were only captured in a subset of sampled sites, we selected bat flies from the six most well-represented bat species for microbiome analysis.

Table 1: Sampling of bats and their corresponding flies used for sequencing. Columns labelled "Total" include female and male flies and flies for which we could not determine a sex due to destruction of the specimen during sampling or transit. We also used flies that had spurious host associations to increase sample size of these species. These are listed under "unknown host."

Bat Family	Bat Species	Bat Sex			Bat Fly Family	Bat Fly Species	Bat Fly Sex		
		F	M	Total			F	M	Total
Phyllostomidae	Artibeus lituratus	36	14	50	Streblidae	Paratrichobius longicrus	16	24	40
	Carollia perspicillata	45	18	65	Streblidae	Paraeuctenodes similis	0	2	2
						Speiseria ambigua	7	6	13
						Strebla guajiro	10	7	18
						Trichobius dugesioides	3	1	4
						Trichobius joblingi	9	12	27
	Desmodus rotundus	9	21	30	Streblidae	Strebla wiedemanni	6	15	21
						Trichobius furmani	1	3	4
	Sturnira lilium	30	6	39	Streblidae	Aspidoptera falcata	14	5	21
						Megistopoda proxima	11	6	17
Vespertilionidae	Myotis nigricans	6	21	27	Nycteribiidae	Basilia andersoni	2	2	4

Bat flies were identified to species morphologically following [31–33]. Access to comparative morphological material was limited, so we barcoded all samples using *cytochrome oxidase I* [34, 35] (Supporting information; NCBI GenBank accession numbers OL847352-OL847639) and confirmed that individual flies identified morphologically as conspecifics belonged to the same genetic clade.

Environmental variables were measured using ArcGIS 10.1 and Fragstats 3.1 using forest cover maps from the Instituto Brasileiro de Geografia and SOS Mata Atlântica (www.sosmataatlantica.com.br; Table S1; [36, 37]. Habitat fragment area (hectares), isolation (shortest distance between a fragment and its nearest neighbor), distance from source (shortest straight line distance from focal point of a fragment to nearest point of REGUA), perimeter-area ratio, proximity index within a 500m and 1000m buffer [36], and forest cover within a 500m and 1000m buffer were calculated. REGUA was treated as the source because it is the largest, most biodiverse patch of forest in the study region. Perimeter-area ratio, proximity index, and forest cover were correlated with habitat patch area, isolation, and distance from source, so only these latter three landscape variables were used for downstream analyses. Area, isolation, and distance from source were $\log_2$ transformed to prevent extremely large or extremely isolated fragments from unduly impacting correlation analyses.

DNA extraction and 16S rRNA metabarcoding

DNA was extracted from 288 bat flies following a wash step and proteinase K digestion using the ZymoBIOMICS DNA Miniprep Kit (Zymo Research, Irvine, CA, USA) in a Biosafety Cabinet, Class 2 (Supporting information). One negative control was used for each extraction kit to control for laboratory and kit contamination. Negative controls were pooled for amplification. Extracted DNA was aliquoted into 96-well plates for amplification of the hypervariable region 4 (V4) of 16S rRNA following well-documented procedures outlined by the Earth Microbiome Project and the Illumina 16S Metagenomic Sequencing Library Preparation guidelines (Supporting information) [38–41]. Of 288 initial libraries, 77 libraries required an additional concentration step to reach the minimum 2nM concentration required for sequencing (Supporting information). Low concentration and high concentration libraries were pooled to reach a concentration of 2nM and sequenced using an Illumina MiSeq v3 Reagent Kit with 2x300bp reads and 18% PhiX spike-in on a MiSeq NGS platform (Illumina, San Diego, CA, USA) at the Bioinformatics and Computational Genomics Laboratory (Hunter College, City University of New York, New York, NY, USA).

De-mulitplexing, Quality Filtering, and Phylogeny Reconstruction

Samples were demultiplexed using the MiSeq Reporter Generate FASTQ workflow. Primer sequences were trimmed from forward and reverse sequence reads using cutadapt v.1.4.2 [42]. Following de-multiplexing, samples were processed using the QIIME2 v.2018.2 pipeline (https://docs.qiime2.org/2018.2) [43–47]. The GreenGenes Database, v.13.5, trimmed to only the 16S rRNA V4 region, was used as a reference to train a naive Bayes q2-feature-classifier for taxonomic identification of amplicon sequence variants (ASVs) [48]. The GreenGenes Database is not able to discern between "*Candidatus* Aschnera," the primary symbiont of some nycteribiid flies [49, 50], and *Arsenophonus*, possibly because the V4 fragment of 16S rRNA does not provide enough resolution. To confirm that ASVs of "*Candidatus* Aschnera chinzeii" were not misidentified, we mapped ASVs identified as *Arsenophonus* and "*Candidatus* Phlomobacter" against reference sequences from the Silva Ribosomal Database (Supporting information; [51–53]. Even though the GreenGenes and Silva Databases distinguish between "*Candidatus* Phlomobacter" and *Arsenophonus*, phylogenetic evidence indicates that "*Candidatus* Phlomobacter" is actually a clade nested within *Arsenophonus* [54, 55].

Contamination is ubiquitous in microbiome studies and especially problematic for low biomass samples [56–58]. To reduce the impact of contaminants, several filtering steps were performed (Supporting information). Briefly, we removed bacteria present in negative controls that were likely contaminants (i.e., low relative abundance and low prevalence in samples); known laboratory contaminants [58]; bacteria classified as mitochondria, chloroplast, or Archaea; bacteria unclassified beyond phylum; and contaminants identified by the R package *decontam* v1.14.0 [60, 61]. Finally, the data were filtered by two coverage depths: 1) all ASVs present at $<0.01\%$ relative abundance within a sample were eliminated from that sample, and 2) all ASVs

present at <0.1% relative abundance were eliminated from that sample. At a minimum within-sample relative abundance of 0.01%, spurious ASVs may remain in the dataset, but at a minimum relative abundance of 0.1%, rare ASVs may be incorrectly excluded [62–64].

As low concentration libraries were used to dilute high concentration libraries prior to sequencing, sequencing effort across samples is not even. To assess the bacterial diversity captured by low concentration libraries compared to high concentration libraries, the *ggrare* function was used from the *phyloseq-extended* suite of tools, which wraps the function *rarefy* from the package *vegan* v.2.5.4 (https://github.com/mahendra-mariadassou/phyloseq-extended/blob/master/R/graphical_methods.R) [65].

Data visualization, Ordination, and PERMANOVA

We constructed bar plots of the relative abundance of bacterial genera in each bat fly species (*ggplot2*, v.3.1.0) [66, 67]. All genera with a relative abundance <1% of the total reads in a bat fly species were condensed into a "Low Abundance" group. Principal coordinates analysis (PCoA), implemented in *phyloseq* v.1.38.0, was used to visualize differences in microbial communities captured by Euclidean distance between *phylogenetic isometric log-ratio*-transformed relative abundances to account for the compositional nature of metabarcoding data (*philr* v.1.20.0, [66–71]; Supporting information).

To test whether landscape variables, parasite variables, or host bat variables were correlated with microbial community composition (i.e., ASV composition and relative abundance), we used PERMANOVA on individual variables and Sequential (Type I) sum of squares on pairs of variables, each with 9,999 permutations (*adonis* command in the R package *vegan*) [72, 73]. Homogeneity of dispersion of each group of microbiomes was confirmed using *betadisper* permuted 999 times with *permutest* (R package *vegan*).

When sampling is uneven, sequential sum of squares is sensitive to the order in which variables appear in the equation. To overcome this limitation, we examined the impact of landscape within the four most well-sampled bat fly species, excluding the southern sites and only considering data filtered using a threshold of 0.01% relative abundance per sample. We ordinated samples within these species separately from the rest of the data and estimated variation explained by landscape variables using PERMANOVA on individual variables.

To examine the impact of habitat patch area and isolation on taxon richness, we constructed boxplots of ASV richness by sampling site, mimicking standard island biogeography plots of richness by area, isolation, and distance to the source. We used a Kruskal-Wallis test to assess whether mean richness was significantly different among sampling sites. Spearman correlation was used to examine ASV richness across continuous ranges of ranked area, isolation, and distance from a source.

Network Reconstruction and Analysis

Bacterial association networks are a way to visualize correlated changes in relative abundance of bacteria across a given sample set. These networks examine microbiomes at the taxon scale, providing a valuable additional analysis to ordination and PERMANOVA, which compare microbiomes at the community scale. While networks provide a useful tool for further examination of microbiomes, the limitations of networks are still being explored and interpretation of networks should be mindful of known caveats (e.g., bias caused by differing sample sizes). Here, we used networks to explore whether the environment or parasite influence microbiome community structure, taking care to examine the role of sample size in network inference. If the environment drives microbiome composition, then we expect that networks will exhibit conserved changes in structure as habitat patches decrease in area or increase in isolation. If parasite species identity impacts microbiome composition, we expect that network structure will be specific to parasite species but will not vary with habitat patch area or isolation.

Networks were reconstructed using *SPIEC-EASI* v.1.1.2 using the Meinshausen and Buhlmann method [74]. Nodes indicate individual bacterial ASVs and edges (i.e., connections between nodes) indicate a linear relationship in the abundances of linked nodes. *SPIEC-EASI* transforms the relative abundance of each ASV using centered log-ratios and then estimates an inverse covariance matrix by solving a regularized

linear regression for each node to determine its conditional independence within the graph [74]. Two nodes are conditionally independent when their abundances are statistically independent given the abundances of all other nodes in the network. The l tuning parameter is used to penalize linear regressions and controls the sparsity of the final network. To select the sparsity of the final network, *SPIEC-EASI* builds graphs from repeated subsamples of the data (i.e., Stability Approach to Regularization Selection, [75] and selects the l value that yields the greatest stability of edge incidences across subsampled graphs. The *SPIEC-EASI* method is distinct from network estimations based on correlation or covariance, which rely on pairwise comparisons and may incorrectly infer an edge between indirectly linked nodes by ignoring the influence of other nodes in the network.

Two types of networks were reconstructed: habitat patch networks estimated from all samples within a single sampling site (1 network per site), and species-specific networks estimated from well-represented parasite species clustered by their occurrence within or outside of REGUA (1 network of within-REGUA samples and 1 network of outside-REGUA samples for each parasite species; Supplemental information). The *lambda.min.ratio* parameter was adjusted until network stability was within 0.002 (habitat patch networks) or 0.003 (species-specific networks) of the target 0.05 threshold, then *nlambda* was set to either 20, 30, or 50 [74, 75]. To examine the impact of sample size on habitat patch networks, we leveraged the subsampling scheme within SpiecEasi to test whether the proportion of highly confident edges (i.e., edges present in at least 80% of sampled networks) was correlated with sample size using the *getOptMerge* function.

Characteristics of networks provide information about the robustness of a community and the function of members of a community [76]. Leading eigenvector modularity of each network and betweenness centrality of each node were estimated in the R package *igraph*, v.1.2.11 [77–80]. Modularity is a measure of the structure of a network, where higher modularity indicates nodes are grouped into tightly interacting neighborhoods with few interactions occurring outside of this neighborhood [81]. Leading eigenvector modularity is an optimization method of community detection with known limitations, despite being widely used [82]. It is unclear how this method performs on sparse, disconnected networks, but modularity estimates may be noisy or difficult to resolve [83]. Betweenness centrality measures the number of times the shortest path between all pairs of nodes in the network travel through a given node, giving an estimate of the influence of a node on the structure of the network [81]. Network connectance is the proportion of realized edges relative to the total possible number of edges [81]. We examined the correlation of modularity, betweenness centrality, and connectance with landscape variables and used Mann-Whitney U and Kruskal-Wallis tests to test for significant differentiation.

Modularity and betweenness centrality are impacted by network size (i.e., number of edges) and shape (degree distribution; [81], making comparisons of summary statistics between networks inaccurate. To account for variation in network size and shape in habitat patch networks, we used several standardization techniques to compare networks: 1) centered modularity compared to mean modularity of patch-specific null distribution; 2) Z-score modularity compared to a patch-specific null distribution; 3) Z-score modularity compared to the mean modularity of measured networks (see Supporting information for detail on null distribution).

As a size-independent method of examining variation between networks, we used the graphlet correlation distance from [84, 85] to ordinate the networks as individual points on a plot using the R packages *pulsar* v.0.3.5 and *orca* v.1.1-1 [86, 87]. The graphlet correlation distance breaks up a network into up-to 4-node graphlets (i.e., all the possible ways that up to 4 nodes can be wired) and counts the number of times each node plays a specific role within a graphlet (i.e., orbit). For example, within a 3-node graphlet where two "leaf" nodes interact with a central node but not each other (i.e., a line of 3 nodes), the "leaf" nodes belong to one orbit and the central node belongs to the second orbit in this graphlet. Using these orbit counts, we estimate the Spearman correlation of orbits across all nodes. Following [88, 89], the Euclidean distance between these correlations can be used to ordinate the networks and more clearly visualize their differentiation. We used k-means clustering to distinguish groups of networks in ordination space.

Results

Of 228 prepared DNA libraries 222 libraries were used for downstream analysis following quality filtering (NCBI SRX13352735-13352962; BioProject PRJNA786937). Filtered libraries ranged in sequencing depth from 2,983 to 66,164 reads. A total of 1,155 ASVs were detected when a 0.01% filtering threshold was applied, while 526 ASVs were found under a 0.1% filtering threshold. Rarefaction curves showed a plateaued asymptote for each library and low concentration libraries fell within the range of ASVs detected in high concentration libraries generated from the same parasite species (Figures S2 and S3).

Composition of sampled bat fly microbiomes

Plots of relative abundance of bacterial genera showed a stark difference between the microbiome communities in the parasite families Nycteribiidae and Streblidae (Figure 2A; Figure S4). While nycteribiid bat flies had high relative abundances of *Wolbachia* and *Bartonella*, streblid bat flies were dominated by *Arsenophonus*. We did not detect "*Candidatus* Aschnera chinzeii" and *Arsenophonus* (including "*Candidatus* Phlomobacter") was detected in low relative abundance in nycteribiid bat flies. Almost no *Wolbachia* was detected in streblid bat flies. *Bartonella* was present in some streblid bat flies, but at much lower relative abundance than in nycteribiid bat flies. *Mycoplasma* was also detected at higher relative abundances in streblid bat flies compared to nycteribiid flies. The flies in southern fragments were dominated by *Wolbachia* and *Bartonella*, likely due to the abundance of nycteribiid flies in these fragments (Figure S5).

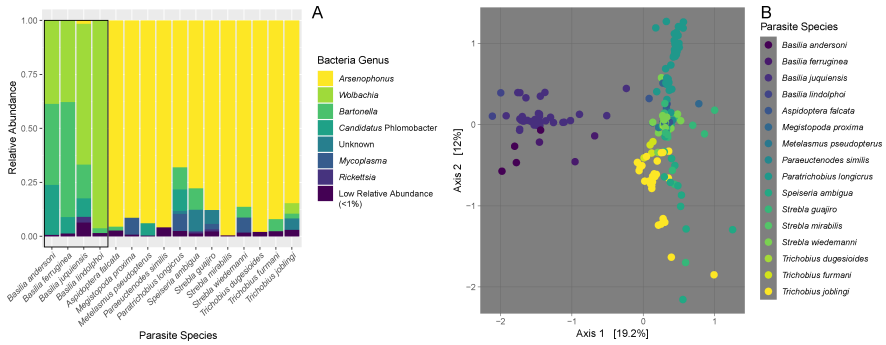


Figure 2: Bat Fly Microbiome Composition– Relative abundance of each bacterial genus summed across repeated samples of each parasite species (A). Colors indicate different bacterial genera and bars represent each parasite species. The black box surrounds the nycteribiid bat flies. Low abundance bacteria were those comprising less than 1% relative abundance in each species. Unknown bacteria could not be identified to genus using the GreenGenes database. Principal Coordinates Analysis on the Euclidean distances between philr-transformed microbial abundances of the complete dataset (B). Colors represent parasite species.

Variation in the microbiome in response to parasite, bat host, and environment

Ordination of microbiomes and PERMANOVA indicated that the parasite (i.e., parasite family and species), the host bat (i.e., bat family, bat sex, and bat individual), and the environment (i.e., region and sampling site) significantly contributed to bat fly microbiome variation (Table 2; Figure 2B and Figure S6). Other variables significantly contributed to microbiome community differentiation (i.e., bat feeding guild, bat species, protection status of sampling site, habitat patch area, and isolation), but violated PERMANOVA's assumption of homoscedasticity. Parasite species, parasite family, bat feeding guild, bat family, bat species, and sampling site had the largest effect sizes, however many of these variables are correlated with each other.

Table 2: Univariate PERMANOVA results indicating the p-value (top; *=p<0.05, **=p<0.01, ***=p<0.0), R2 (middle) and p-value for homoscedasticity (bottom, significance indicates violation of the assumptions of PERMANOVA). Results are provided for the full dataset (all sites at 0.01% threshold for bacterial relative abundance per sample), the strictly filtered dataset (all sites at 0.1% threshold for bacterial relative abundance), the samples collected from REGUA area sites (excluding southern sites), and the samples collected from fragments outside of REGUA (excluding southern sites). Grey shading indicates variables that significantly differentiated microbiomes and did not violate the assumptions of homoscedasticity.

		All Localities	All Localities, Strictly Filtered	REGUA Area	REGUA Area Unprotected
Parasite Variables	Parasite Family	0.0***	0.0***	0.0***	0.0***
		0.18	0.22	0.15	0.13
		0.02*	0.06	0.04*	0.09
	Parasite Species	0.0***	0.0***	0.0***	0.0***
		0.5	0.5	0.5	0.52
		0.12	0.13	0.08	0.06
	Parasite Sex	0.39	0.29	0.08	0.15
		0.01	0.01	0.01	0.01
		0.05*	0.01*	0.0**	0.0**
Bat Variables	Bat Feeding Guild	0.0***	0.0***	0.0***	0.0***
		0.18	0.21	0.15	0.17
		0.0**	0.0**	0.0**	0.01*
	Bat Family	0.0***	0.0****	0.0***	0.0***
		0.16	0.2	0.12	0.13
		0.02*	0.09	0.04*	0.09
	Bat Species	0.0***	0.0***	0.0***	0.0***
		0.27	0.28	0.26	0.28
		0.0**	0.0**	0.0**	0.0**
	Bat Sex	0.0***	0.0***	0.03*	0.22
		0.02	0.02	0.01	0.01
		0.62	0.14	0.87	0.39
	Individual Bat	0.0**	0.0**	0.0***	0.04*
		0.01	0.01	0.02	0.02
		0.34	0.46	0.32	0.5
Landscape Variables	Region	0.0***	0.0***		
		0.05	0.05		
		0.47	0.33		
	Protec-			0.0***	

Sequential sum-of-squares with free permutation indicated that parasite species significantly impacted microbiome community structure, consistent with single variable PERMANOVA (Table 3). Habitat patch protection status and region also significantly contributed to microbiome variation. In the four most well-sampled species, none of the test variables significantly explained microbiome variation without violating PERMANOVA assumptions (Table S2; Figure S7).

Table 3: Sequential Sum-of-Squares where the impact of each variable is considered after the impact of parasite species is accounted. Grey shading indicates significant impact without violating assumptions of PERMANOVA. The PERMANOVA p-value (top; *=p<0.05, **=p<0.01, ***=p<0.0), R2 (middle) and p-value for homoscedasticity (bottom, significance indicates violation of the assumptions of PERMANOVA) are provided for each dataset.

	All Localities	All Localities, Strictly Filtered	REGUA Area	REGUA Area Unprotected
Parasite Species	0.0***	0.0***	0.0***	0.0***
	0.5	0.5	0.48	0.53
	0.11	0.14	0.08	0.06
Bat Species	0.52	0.52	0.54	0.58
	0.01	0.01	0.02	0.02
	0.0**	0.01**	0.0***	0.0**
Bat Sex	0.16	0.23	0.22	0.33
	0.0	0.0	0.0	0.0
	0.61	0.14	0.86	0.41
Individual Bat	0.22	0.24	0.08	0.66
	0.0	0.0	0.0	0.0
	0.33	0.46	0.28	0.52
Region	0.01**	0.01*		
	0.01	0.01		
	0.46	0.32		
Log_2 **Area**			0.03*	0.53
			0.01	0.0
			0.03*	0.11
Log_2 **Isolation**			0.23	0.29
			0.0	0.01
			0.04*	0.11
Log_2 **Distance to Source**				0.4
				0.0
				0.12
Protection Status			0.02*	
			0.01	
			0.06	
Sampling Site				0.05
				0.05
				0.09

Table 3: Sequential Sum-of-Squares where the impact of each variable is considered after the impact of parasite species is accounted. Grey shading indicates significant impact without violating assumptions of PERMANOVA. The PERMANOVA p-value (top; *=p<0.05, **=p<0.005, ***=p<0.0005), R2 (middle) and p-value for homoscedasticity (bottom, significance indicates violation of the assumptions of PERMANOVA) are provided for each dataset.

Bacterial taxon richness across habitat patches

While the three sampled sites within REGUA had the highest ASV richness, there was no pattern of decreasing ASV richness with decreasing area, increasing isolation, or increasing distance from a source (Spearman; Area: rho=0.0613, p-value=0.4031; Isolation: rho=0.0020, p-value=0.9787; Distance to Source: rho=-0.0564, p-value=0.4421; Figure 3). Median bacterial ASV richness fell between 6 and 11 for each sampled parasite individual, but the range of ASV richness per fragment varied dramatically. There was no significant difference between sampling sites in mean ASV richness (Kruskal-Wallis chi-squared statistic=17.856, p-value=0.1201).

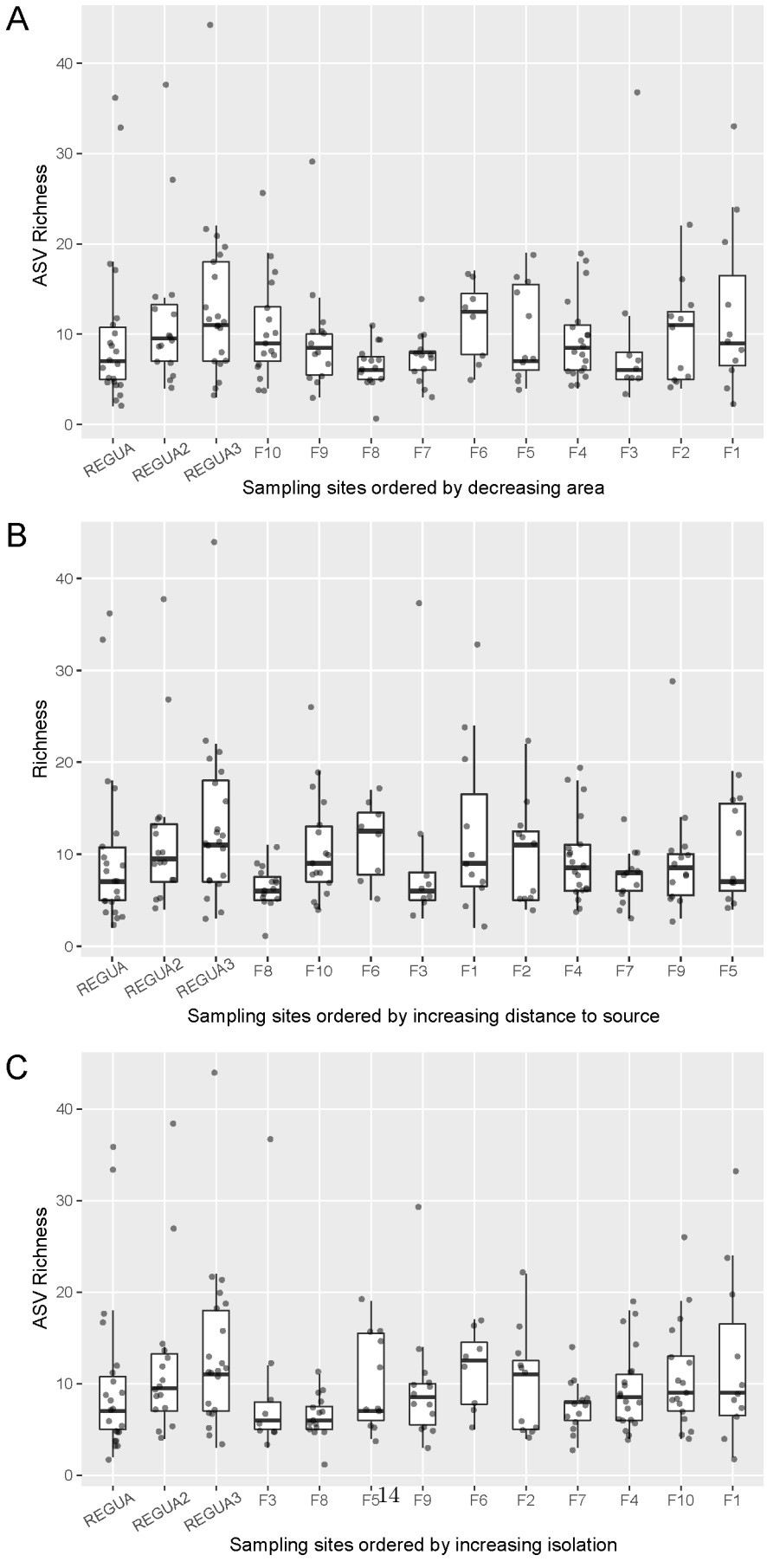


Figure 3: Taxon Richness by Site– Box and whisker plot of the bacterial ASV richness in each sampling site in order of decreasing area (A), increasing distance from a source (B), and increasing isolation (C).

Bacterial association networks

The connectance and betweenness centrality of habitat patch networks varied with habitat patch effects, but modularity did not. Network connectance was significantly lower in REGUA patches than in networks constructed for sites outside of REGUA (Figure 5B; Wilcoxon Rank Sum test, p-value =0.012). Betweenness centrality was higher in networks for patch F4, F2, and F1 (i.e., smallest patches), corresponding with their band-like network structure (Figure 4), and centrality was significantly different between networks from large and small patches (Figure 5c; Kruskal-Wallis chi-squared statistic = 300.23, p-value $< 2.2 \times 10^{-16}$). Raw and Z-score modularity of habitat patch networks did not significantly differ between REGUA and non-REGUA sites, but tended to be lower in patches outside REGUA (Figure S8). Z-score modularity using the measured networks for standardization did not control for network size and shape, and mimicked the pattern exhibited by raw modularity. Modularity measures standardized by null distributions did not support patterns of decreasing modularity with decreasing patch area (Figure S8).

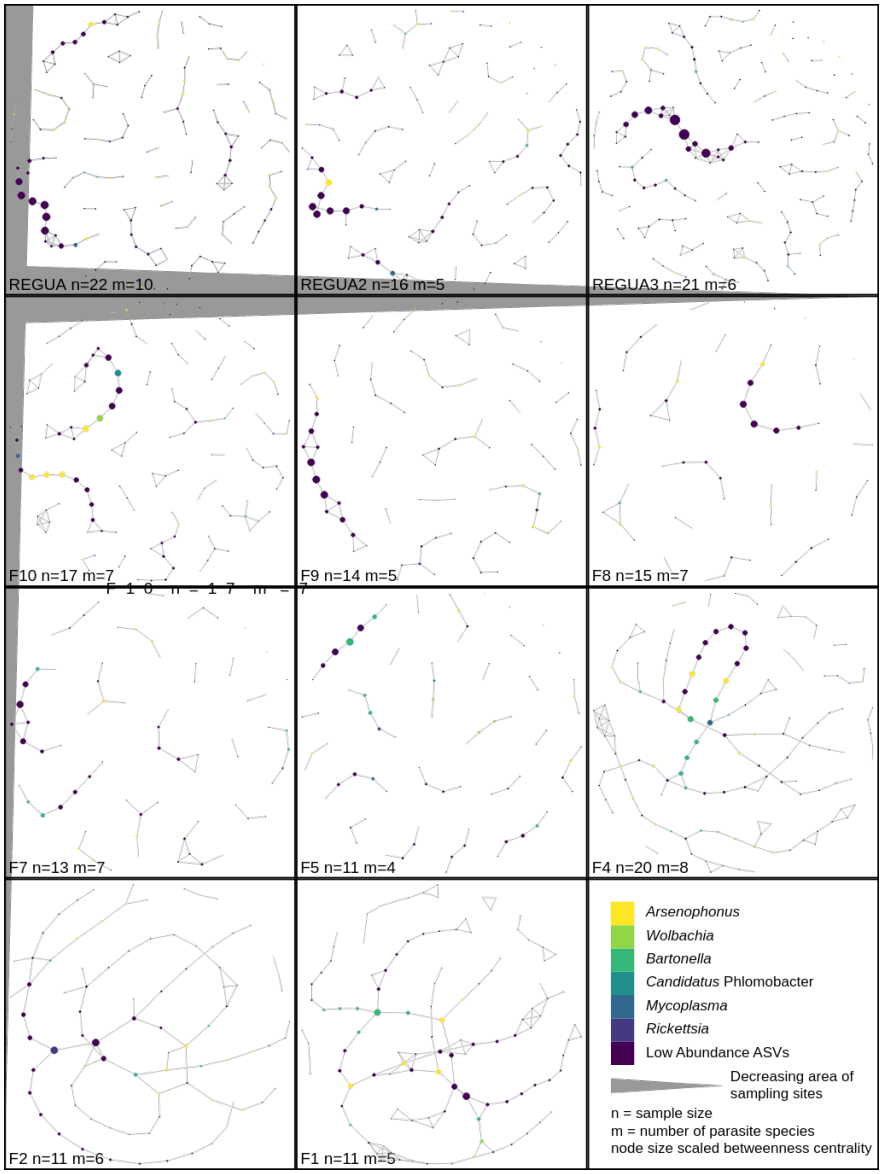


Figure 4: Bacterial Association Networks– Networks for each sampling site are ordered by decreasing habitat patch area, with the largest sites in the top left and the smallest sites in the bottom right. The size of the nodes in the networks corresponds to the z-score betweenness centrality of that node scaled by the range of betweenness centralities detected within the network.

In habitat patch networks reconstructed for the three REGUA sites and large patches, low-abundance bacteria had higher betweenness centrality than other bacteria in these networks, acting to connect graph neighborhoods (Figure 4). As habitat patch networks became less modular with decreasing habitat patch area, nodes with high betweenness centrality were typically low-abundance bacteria and bacteria in the genera *Arsenophonus* (including *Candidatus* Phlomobacter), *Wolbachia*, and *Bartonella*.

While there was no impact of sample size (i.e., number of parasite individuals) on ASV richness in each network, a greater number of samples may allow detection of more edges between nodes and change the size of a network. Sample size was lowest in small fragments and highest in large fragments, with the

exception of F4 which has intermediate area, isolation, and distance to source measurements, but supports a high diversity of parasites. We did not detect a significant correlation between sample size and network modularity (Spearman rho=0.3945, p-value=0.229). Using the subsampling scheme within SpiecEasi, we did not detect a correlation of sample size and proportion of high-confidence network edges (Spearman rho=0.0275, p-value=0.936).

Using ordinations of graphlet orbits as a size-independent comparison of networks [84, 88, 89], habitat patch networks vary with habitat area (Figure 5D) and species-specific networks are distinct within and outside of REGUA (Figure S8). Principal coordinate axis 2 corresponded well to decreasing habitat area, with large fragments positioned higher on the access and gradually decreasing in patch area lower on the axis. Principal coordinate axis 1 primarily illustrated variation in networks from patches F8 and F5. This variation does not correspond with environment, parasite, or bat variables. K-means clustering with 3 groups separated large habitat patch networks (REGUA networks, F10, F9, F7) from small patch networks (F4, F2, F1), and F8 and F5 formed a unique cluster. Species-specific networks also indicated distinctions between within-REGUA and outside-REGUA networks. All within-REGUA networks occupied unique ordination space from outside-REGUA networks within species.

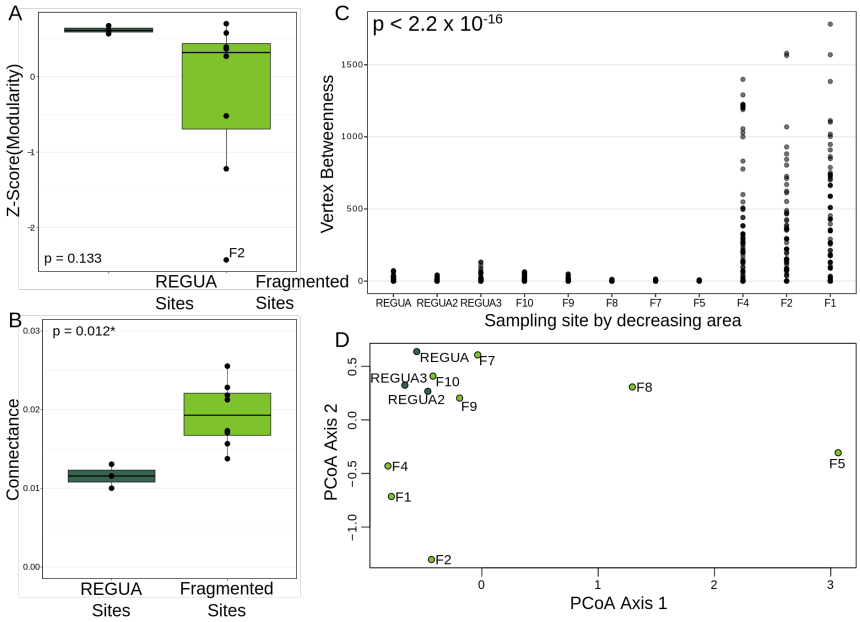


Figure 5: Summary Statistics of Network Plots– Z-score modularity of habitat patch networks between REGUA and outside-REGUA sites (A), network connectance of habitat patch networks in REGUA and outside-REGUA sites (B), vertex betweenness centrality by fragment decreasing by habitat patch area (C), principal coordinates analysis of the distribution of network orbits for habitat patch networks (D). Dark green corresponds to REGUA sites and lime green corresponds to sites outside of REGUA for which networks were reconstructed.

Discussion

Our research builds upon previous evidence that the environment influences microbiome composition in addition to host factors [90–93]. However, previous studies have been primarily conducted on the microbiomes of free-living hosts or microbiomes within a single host species, preventing the examination of

how the environment might impact communities of interacting macro- and microorganisms. Contrary to expectations, habitat loss did not lead to a decrease in bacterial diversity. Instead, habitat loss impacted bacterial community structure, which includes diversity and relative abundance. The interactions of bat flies with the broader environment are filtered through their obligate associations with host bats, yet the signal of environmental change is also detected in the composition of bat fly microbiomes. This indicates that environmental degradation may have cascading consequences through hierarchical communities of interacting organisms.

Parasite and environment as drivers of microbiome composition

Parasite species identity is the strongest predictor of microbiome composition (Tables 2 and 3; Figure 2B). Specificity of arthropod microbiomes has been previously found in tsetse flies (Glossinidae; [94], which are also members of the Hippoboscoidea. That microbiome composition shows such a strong signal of parasite species identity indicates that either more of the microbiome is maternally inherited than previously understood, or that even non-maternally inherited bacteria may be maintained through life history traits (e.g., host bat associations, microclimatic preference; horizontal transmission of bacteria).

Habitat patch area and protection status, but not degree of isolation or distance to source, have a measurable effect on the microbiome of bat flies, but less explanatory power than parasite species (Tables 2 and 3; Figures 3, 4, 5). While bacterial ASV richness does not vary following expectations of island biogeography theory (Figure 3), examining both relative abundance and diversity of bacteria using PERMANOVA and bacterial association networks provided a clear statistical signal that habitat patch area (measured as area and protection status) is correlated with microbiome composition (Table 3; Figure 5). Other measures like isolation and distance to source had no impact on bat, parasite, or microbiome communities, likely because bats can easily move between these patches. However, even when habitat patches are proximal, fragmented landscapes in the Atlantic Forest have been previously shown to significantly impact bat-plant and bat-ectoparasite interaction networks [95].

Even though changes in the bat and parasite communities explain most of the variation in microbiome association networks, the environment may also directly impact bat fly microbiome communities. The structure of habitat patch networks covaried with habitat area and parasite species-specific networks were consistently different within REGUA compared to outside REGUA. In addition, parasite and bat species diversity did not always correspond to a specific network structure. REGUA and fragments F10 (the largest fragment outside of REGUA, 9 bat fly species) and F4 (8 bat fly species) have the greatest parasite richness, but F4 is intermediate in area and isolation. Networks from patch F4 consistently cluster with other networks from small patches (Figure 5C,D), despite having similar parasite species richness to large fragments. That patch F4 has high parasite species richness but similar bacterial association networks to smaller fragments (i.e., lower parasite species richness), suggests that bat and parasite community composition does not solely explain microbiome variation. The environment may also directly drive composition of parasite-associated microbiomes.

Implications of changes in bacterial network structure in response to environment

Bacteria with high betweenness centrality may act as hub species that maintain the stability of a network [19, 92, 93]. In small fragments, *Arsenophonus*, *Wolbachia*, and *Bartonella* had high betweenness centrality, but these bacterial taxa were less central to the networks from large fragments despite maintaining high relative abundance in flies at these sites. Decreasing betweenness centrality may be indicative of changing interactions between bacteria in response to environmental perturbations. Bacteria of blood-feeding insects play an important role in vector competence in insects [19, 96, 97]. For example, in tsetse flies, primary bacterial endosymbionts in the genus *Wigglesworthia* impede the invasion of trypanosome parasites by assisting host defenses and subsequently decrease the competence of tsetse flies to vector these harmful parasites to downstream hosts including humans [98]. As bat flies are important arthropod vectors of bat pathogens [99], changes in the structure of their microbiomes in response to habitat loss may have implications for the disease ecology of arthropod vectored pathogens in bats.

Modules may delimit groups of bacteria with specific functional specializations and/or groups that respond in similar ways to environmental variables [99, 100]. Higher modularity may protect a community of free-living organisms from invading pathogens, because a pathogen would be isolated to one module within the community (i.e., diversity-stability debate) [100, 101]. This hypothesis may be applicable to bacterial networks if pathogens are limited in transmission by direct competition with endogenous bacteria. However, high modularity in bacterial association networks may also reflect the absence of microbiome-mediated host defenses against pathogen invasion. If we revisit the tsetse fly example, *Wigglesworthia* and trypanosome parasites would share an edge in a microbe association network because *Wigglesworthia* abundance and prevalence is associated with low trypanosome abundance and prevalence mediated by host defenses. The absence of this interaction in a network might indicate that *Wigglesworthia*-induced host response to trypanosome invasion is impaired by other microbes in the network (e.g., priority effects) or other aspects of host health. Trypanosome transmission would be less regulated in this instance, but modularity would be high as long as trypanosome abundance and prevalence did not lead to changes in the abundance or prevalence of other microbes. In the case of bat flies, the consequences of more isolated modules for the functionality and stability of the microbiome are unclear and merit future investigation.

The missing primary symbionts of Brazilian nycteribiid flies

Arsenophonus has been previously identified as a primary symbiont of streblid bat flies and the closely-related "*Candidatus* Aschnera chinzeii" is the hypothesized primary symbiont of some nycteribiid bat flies [49, 50, 102–104]. The high relative abundance and prevalence of *Arsenophonus* in all streblid bat flies sampled in our study is consistent with the hypothesis that *Arsenophonus* acts as the primary symbiont in the family Streblidae. However, it is unlikely that *Arsenophonus* or "*Candidatus* Aschnera chinzeii" are acting as the primary symbionts of the nycteribiid species that we sampled (Figure 2A). Previous studies that have identified "*Candidatus* Aschnera chinzeii" as the primary symbiont in nycteribiid flies examined species that are geographically limited to Africa, Asia, Europe, and Oceania [102, 103]. Species within *Basilia*, the only globally distributed bat fly genus [105, 106], have varying symbiont associations [49, 50, 103, 104]. Only one previous study has examined symbionts of *Basilia* from The Americas [107]. This study detected an *Arsenophonus* variant in two individuals of *Basilia boardmani* (Nycteribiinae; restricted to North America) that was distinct from "*Candidatus* Aschnera chinzeii" and from *Arsenophonus* detected in other nyteribiid species. In the *Basilia* species sampled from The Atlantic Forest, of which 3 are limited to South America (*B. andersoni*, *B. juquiensis*, *B. lindolphoi*) and 1 is found in North and South America (*B. ferruginea*), we detect no "*Candidatus* Aschnera chinzeii" and low relative abundance or an absence of *Arsenophonus* (including "*Candidatus* Phlomobacter"). These findings in combination with previous studies suggest that neither "*Candidatus* Aschnera chinzeii" nor *Arsenophonus* act as the primary symbiont of the sampled *Basilia* species. It may be that *Wolbachia* and *Bartonella* both have high relative abundance in nycteribiid bat flies because they are acting as primary or facultative symbionts, or because they are acting as a reproductive parasite and pathogen, respectively [107–113]. If the latter is true, the primary symbiont may be one of the less abundant bacteria or another microbe not detected by 16S rRNA sequencing (e.g. fungi; [114].

Summation

Understanding of community structure and function must be extended beyond an exclusively macroorganismal view to include the layers of microorganisms that also participate in defining an ecological community. By examining the hierarchical interactions between bats, bat flies, and bat fly bacterial microbiomes within a largely deforested landscape, we attempted to more accurately characterize the consequences of environmental change for a wildlife community. Reduced fragment area led to generally less diverse and less abundant bat fly communities, which led to less modular bacterial association networks, but without a decrease in bacterial ASV richness. Our data highlight the importance of considering ecological responses of microorganismal taxa. Future research is needed in order to capture the full impact of continuing deforestation and habitat change on ecological communities.

Acknowledgements

Thank you to REGUA staff for their assistance in the field and Dr. Marcelo Weksler and Museu Nacional/UFRJ staff for logistical help. This research was funded by Richard Gilder Graduate School Dissertation Research Fellowships to KAS and MRI, and a CAPES "science without borders" Fellowship to TSMT with additional funding from Queen Mary University of London to ELC and from The American Museum of Natural History to SLP. This research was also enabled by funding to ELC from the Natural Sciences and Engineering Research Council of Canada through the Discovery Grants Program, support from the Government of Canada's New Frontiers in Research Fund (NFRFT-2020-0073) and by support from Genome Canada and Ontario Genomics to BIOSCAN–Canada through the Large Scale Applied Research Program. All field work was conducted under Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis permit 19037-1.

Conflict of Interest Statement

The authors have no competing interests to declare.

Data Accessibility Statement

Raw 16S rRNA reads are available under NCBI SRA accessions SRX13352735-13352962 and BioProject PRJNA786937, COI sequences of bat flies are available on GenBank (OL847352-OL847639), metadata and all code used for analyses is available on Github (DOI: 10.5281/zenodo.6796123).

Literature Cited

1. Alroy J. Effects of habitat disturbance on tropical forest biodiversity. *Proc Natl Acad Sci U S A* 2017; **114**: 6056–6061.

2. Gatti LV, Basso LS, Miller JB, Gloor M, Gatti Domingues L, Cassol HLG, et al. Amazonia as a carbon source linked to deforestation and climate change. *Nature* 2021; **595**: 388–393.

3. Ellwanger JH, Kulmann-Leal B, Kaminski VL, Valverde-Villegas JM, Veiga ABGDA, Spilki FR, et al. Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. *An Acad Bras Cienc* 2020; **92**: e20191375.

4. Morand S, Lajaunie C. Outbreaks of vector-borne and zoonotic diseases are associated with changes in forest cover and oil palm expansion at global scale. *Front Vet Sci* 2021; **8**: 661063.

5. Keenan RJ, Reams GA, Achard F, de Freitas JV, Grainger A, Lindquist E. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *For Ecol Manage* 2015; **352**: 9–20.

6. Rezende CL, Scarano FR, Assad ED, Joly CA, Metzger JP, Strassburg BBN, et al. From hotspot to hopespot: An opportunity for the Brazilian Atlantic Forest. *Perspectives in Ecology and Conservation* 2018; **16**: 208–214.

7. Yarwood SA. The role of wetland microorganisms in plant-litter decomposition and soil organic matter formation: a critical review. *FEMS Microbiol Ecol* 2018; **94**.

8. Kock RA, Orynbayev M, Robinson S, Zuther S, Singh NJ, Beauvais W, et al. Saigas on the brink: Multidisciplinary analysis of the factors influencing mass mortality events. *Sci Adv* 2018; **4**: eaao2314.

9. Murdock CC, Blanford S, Hughes GL, Rasgon JL, Thomas MB. Temperature alters *Plasmodium* blocking by *Wolbachia*. *Sci Rep* 2014; **4**: 3932.

10. MacArthur RH, Wilson EO. An equilibrium theory of insular zoogeography. *Evolution* 1963; **17**: 373–387.

11. Krasnov BR, Shenbrot GI, Medvedev SG. Host–habitat relations as an important determinant of spatial distribution of flea assemblages (Siphonaptera) on rodents in the Negev Desert. *Parasitology* 1997.

12. Poulin R. Are there general laws in parasite ecology? *Parasitology* 2007.

13. Speer KA, Dheilly NM, Perkins SL. Microbiomes are integral to conservation of parasitic arthropods. *Biol Conserv* 2020; 108695.

14. Bell T, Ager D, Song J-I, Newman JA, Thompson IP, Lilley AK, et al. Larger islands house more bacterial taxa. *Science* 2005; **308**: 1884.

15. Zinger L, Boetius A, Ramette A. Bacterial taxa-area and distance-decay relationships in marine environments. *Mol Ecol* 2014; **23**: 954–964.

16. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* 2006; **4**: 102–112.

17. Carbonero F, Oakley BB, Purdy KJ. Metabolic flexibility as a major predictor of spatial distribution in microbial communities. *PLoS One* 2014; **9**: e85105.

18. van der Gast CJ. Microbial biogeography: the end of the ubiquitous dispersal hypothesis? *Environ Microbiol* . 2015. , **17**: 544–546

19. Weiss B, Aksoy S. Microbiome influences on insect host vector competence. *Trends Parasitol* 2011; **27**: 514–522.

20. Gupta A, Nair S. Dynamics of insect-microbiome interaction influence host and microbial symbiont. *Front Microbiol* 2020; **11**: 1357.

21. Dick CW, Dittmar K. Parasitic bat Flies (Diptera: Streblidae and Nycteribiidae): Host specificity and potential as vectors. In: Klimpel S, Mehlhorn H (eds). *Bats (Chiroptera) as Vectors of Diseases and Parasites*. 2014. Springer, Berlin, Heidelberg, pp 131–155.

22. Speer KA, Luetke E, Bush E, Sheth B, Gerace A, Quicksall Z, et al. A fly on the cave wall: Parasite genetics reveal fine-scale dispersal patterns of bats. 2019.

23. Patterson BD, Dick CW, Dittmar K. Sex biases in parasitism of neotropical bats by bat flies (Diptera: Streblidae). *J Trop Ecol* 2008; **24**: 387–396.

24. Hiller T, Brändel SD, Honner B, Page RA, Tschapka M. Parasitization of bats by bat flies (Streblidae) in fragmented habitats. *Biotropica* 2020; **72**: 617.

25. Kikuchi Y, Tada A, Musolin DL, Hari N, Hosokawa T, Fujisaki K, et al. Collapse of insect gut symbiosis under simulated climate change. *MBio* 2016; **7**.

26. Thapa S, Zhang Y, Allen MS. Effects of temperature on bacterial microbiome composition in *Ixodes scapularis* ticks. *Microbiologyopen* 2019; **8**: e00719.

27. Souto Martins Teixeira T. Bats in a fragmented world. 2019. Queen Mary University of London.

28. Emmons L, Feer F. Neotropical rainforest mammals: a field guide. 1997. sidalc.net.

29. Reis NR, Fregonezi MN, Peracchi AL, Shibatta OA. Morcegos do Brasil: guia de campo. 2013. Technical Books Editora.

30. Sikes RS, Care A, of Mammalogists UC of TAS. 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *J Mammal* 2016; **97**: 663–688.

31. Wenzel RL. The streblid batflies of Venezuela (Diptera: Streblidae). *Brigham Young University Science Bulletin, Biological Series* 1976; **20**: 1.

32. Graciolli G, de Carvalho CJB. Moscas ectoparasitas (Diptera, Hippoboscoidea) de morcegos (Mammalia, Chiroptera) do Estado do Paraná. 11. Streblidae. Chave pictórica para gêneros e espécies 1. *RevIa bras Zool* 2001; **18**: 907–960.

33. Graciolli G, de Carvalho CJB. Moscas ectoparasitas (Diptera, Hippoboscoidea, Nycteribiidae) de morcegos (Mammalia, Chiroptera) do Estado do Paraná, Brasil. I. Basilia, taxonomia e chave pictórica para as espécies 1. *RevIa bras Zool* 2001; **18**: 33–49.

34. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 1994; **3**: 294–299.

35. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B: Biological Sciences* 2003; **270**: 313–321.

36. Gustafson EJ, Parker GR. Relationships between landcover proportion and indices of landscape spatial pattern. *Landsc Ecol* 1992; **7**: 101–110.

37. McGarigal K, Cushman SA, Neel MC, Ene E. FRAGSTATS: spatial pattern analysis program for categorical maps. 2002.

38. Gilbert JA, Meyer F, Antonopoulos D, Balaji P, Brown CT, Brown CT, et al. Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project. *Stand Genomic Sci* 2010; **3**: 243–248.

39. Gilbert JA, Jansson JK, Knight R. The Earth Microbiome project: successes and aspirations. *BMC Biol* 2014; **12**: 69.

40. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol* 2015; **75**: 129–137.

41. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 2016; **18**: 1403–1414.

42. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011; **17**: 10–12.

43. Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002; **30**: 3059–3066.

44. Katoh K, Kuma K-I, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 2005; **33**: 511–518.

45. Katoh K, Toh H. PartTree: an algorithm to build an approximate tree from a large number of unaligned sequences. *Bioinformatics* 2007; **23**: 372–374.

46. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 2010; **5**: e9490.

47. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; **13**: 581–583.

48. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006; **72**: 5069–5072.

49. Hosokawa T, Nikoh N, Koga R, Satô M, Tanahashi M, Meng X-Y, et al. Reductive genome evolution, host–symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies. *ISME J* 2012; **6**: 577–587.

50. Duron O, Schneppat UE, Berthomieu A, Goodman SM, Droz B, Paupy C, et al. Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies. *Mol Ecol* 2014; **23**: 2105–2117.

51. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013; **41**: D590-6.

52. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* 2014; **42**: D643-8.

53. Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, et al. 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J Biotechnol* 2017; **261**: 169–176.

54. Nováková E, Hypsa V, Moran NA. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiol* 2009; **9**: 143.

55. Bressan A, Terlizzi F, Credi R. Independent origins of vectored plant pathogenic bacteria from arthropod-associated *Arsenophonus* endosymbionts. *Microb Ecol* 2012; **63**: 628–638.

56. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014; **12**: 87.

57. Weiss S, Amir A, Hyde ER, Metcalf JL, Song SJ, Knight R. Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol* 2014; **15**: 564.

58. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in low microbial biomass microbiome studies: Issues and recommendations. *Trends Microbiol* 2019; **27**: 105–117.

59. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 2018; **6**: 226.

60. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* 2013; **10**: 57–59.

61. Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods Ecol Evol* 2018; **9**: 134–147.

62. McMurdie PJ, Holmes S. Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. *Pac Symp Biocomput* 2012; 235–246.

63. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013; **8**: e61217.

64. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: community ecology package. R package version 2.4–4. 2017. 2018.

65. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016. Springer.

66. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* 2014; **2**: 15.

67. Gloor GB, Reid G. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can J Microbiol* 2016; **62**: 692–703.

68. Tsilimigras MCB, Fodor AA. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Ann Epidemiol* 2016; **26**: 330–335.

69. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: And this is not optional. *Front Microbiol* 2017; **8**: 2224.

70. Silverman JD, Washburne AD, Mukherjee S, David LA. A phylogenetic transform enhances analysis of compositional microbiota data. *Elife* 2017; **6**.

71. Xia Y, Sun J. Hypothesis testing and statistical analysis of microbiome. *Genes Dis* 2017; **4**: 138–148.

72. Anderson MJ, Walsh DCI. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecol Monogr* 2013; **83**: 557–574.

73. Anderson MJ. Permutational Multivariate Analysis of Variance (PERMANOVA). In: Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F, Teugels JL (eds). *Wiley StatsRef: Statistics Reference Online*. 2014. John Wiley & Sons, Ltd, Chichester, UK, pp 1–15.

74. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol* 2015; **11**: e1004226.

75. Liu H, Roeder K, Wasserman L. Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models. *Adv Neural Inf Process Syst* 2010; **24**: 1432–1440.

76. Röttjers L, Faust K. From hairballs to hypotheses-biological insights from microbial networks. *FEMS Microbiol Rev* 2018; **42**: 761–780.

77. Freeman LC. Centrality in social networks conceptual clarification. *Soc Networks* 1978; **1**: 215–239.

78. Brandes U. A faster algorithm for betweenness centrality. *J Math Sociol* 2001; **25**: 163–177.

79. Csardi G, Nepusz T, Others. The igraph software package for complex network research. *InterJournal, Complex Systems* 2006; **1695**: 1–9.

80. Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. *Phys Rev E Stat Nonlin Soft Matter Phys* 2006; **74**: 036104.

81. Delmas E, Besson M, Brice M-H, Burkle LA, Dalla Riva GV, Fortin M-J, et al. Analysing ecological networks of species interactions: Analyzing ecological networks. *Biol Rev* 2019; **94**: 16–36.

82. Fortunato S, Hric D. Community detection in networks: A user guide. *arXiv [physics.soc-ph]* . 2016.

83. Singh A, Humphries MD. Finding communities in sparse networks. *Sci Rep* 2015; **5**: 8828.

84. Yaveroğlu ÖN, Malod-Dognin N, Davis D, Levnajic Z, Janjic V, Karapandza R, et al. Revealing the hidden language of complex networks. *Sci Rep* 2014; **4**: 4547.

85. Przulj N. Biological network comparison using graphlet degree distribution. *Bioinformatics* 2007; **23**: e177-83.

86. Hočevar T, Demšar J, Others. Computation of graphlet orbits for nodes and edges in sparse graphs. *J Stat Softw* 2016; **71**.

87. Müller CL, Bonneau R, Kurtz Z. Generalized stability approach for regularized graphical models. *arXiv [statME]* . 2016.

88. Mahana D, Trent CM, Kurtz ZD, Bokulich NA, Battaglia T, Chung J, et al. Antibiotic perturbation of the murine gut microbiome enhances the adiposity, insulin resistance, and liver disease associated with high-fat diet. *Genome Med* 2016; **8**: 48.

89. Ruiz VE, Battaglia T, Kurtz ZD, Bijnens L, Ou A, Engstrand I, et al. A single early-in-life macrolide course has lasting effects on murine microbial network topology and immunity. *Nat Commun* 2017; **8**: 518.

90. Amato KR, Yeoman CJ, Kent A, Righini N, Carbonero F, Estrada A, et al. Habitat degradation impacts black howler monkey (*Alouatta pigra*) gastrointestinal microbiomes. *ISME J* 2013; **7**: 1344–1353.

91. Avena CV, Parfrey LW, Leff JW, Archer HM, Frick WF, Langwig KE, et al. Deconstructing the bat skin microbiome: Influences of the host and the environment. *Front Microbiol* 2016; **7**: 1753.

92. Becker CG, Longo AV, Haddad CFB, Zamudio KR. Land cover and forest connectivity alter the interactions among host, pathogen and skin microbiome. *Proc Biol Sci* 2017; **284**.

93. Ingala MR, Becker DJ, Bak Holm J, Kristiansen K, Simmons NB. Habitat fragmentation is associated with dietary shifts and microbiota variability in common vampire bats. *Ecol Evol* 2019; **9**: 6508–6523.

94. Aksoy E, Telleria EL, Echodu R, Wu Y, Okedi LM, Weiss BL, et al. Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota. *Appl Environ Microbiol* 2014; **80**: 4301–4312.

95. Mello RM, Laurindo RS, Silva LC, Pyles MV, Mancini MCS, Dáttilo W, et al. Landscape configuration and composition shape mutualistic and antagonistic interactions among plants, bats, and ectoparasites in human-dominated tropical rainforests. *Acta Oecol* 2021; **112**: 103769.

96. Cirimotich CM, Ramirez JL, Dimopoulos G. Native microbiota shape insect vector competence for human pathogens. *Cell Host Microbe* 2011; **10**: 307–310.

97. Sassera D, Epis S, Pajoro M, Bandi C. Microbial symbiosis and the control of vector-borne pathogens in tsetse flies, human lice, and triatomine bugs. *Pathog Glob Health* 2013; **107**: 285–292.

98. Weiss BL, Wang J, Maltz MA, Wu Y, Aksoy S. Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers. *PLoS Pathog* 2013; **9**: e1003318.

99. Obame-Nkoghe J, Rahola N, Bourgarel M, Yangari P, Prugnolle F, Maganga GD, et al. Bat flies (Diptera: Nycteribiidae and Streblidae) infesting cave-dwelling bats in Gabon: diversity, dynamics and potential role in *Polychromophilus melanipherus* transmission. *Parasit Vectors* 2016; **9**: 333.

100. Krause AE, Frank KA, Mason DM, Ulanowicz RE, Taylor WW. Compartments revealed in food-web structure. *Nature* 2003; **426**: 282–285.

101. Stouffer DB, Bascompte J. Understanding food-web persistence from local to global scales. *Ecol Lett* 2010; **13**: 154–161.

102. Trowbridge RE, Dittmar K, Whiting MF. Identification and phylogenetic analysis of *Arsenophonus*- and *Photorhabdus*-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). *J Invertebr Pathol* 2006; **91**: 64–68.

103. Morse SF, Bush SE, Patterson BD, Dick CW, Gruwell ME, Dittmar K. Evolution, multiple acquisition, and localization of endosymbionts in bat flies (Diptera: Hippoboscoidea: Streblidae and Nycteribiidae). *Appl Environ Microbiol* 2013; **79**: 2952–2961.

104. Wilkinson DA, Duron O, Cordonin C, Gomard Y, Ramasindrazana B, Mavingui P, et al. The bacteriome of bat flies (Nycteribiidae) from the Malagasy Region: a community shaped by host ecology, bacterial transmission mode, and host-vector specificity. *Appl Environ Microbiol* 2016; **82**: 1778–1788.

105. Graciolli G, Dick CW. Checklist of World Nycteribiidae (Diptera: Hippoboscoidea). https://www.researchgate.net/publication/322579074_CHECKLIST_OF_WORLD_NYCTERIBIIDAE_DIPTERA_HIPPOBOSCOIDEA. .

106. Graciolli G, Dick CW. Checklist of World Streblidae (Diptera: Hippoboscoidea). https://www.researchgate.net/publication/322578987_CHECKLIST_OF_WORLD_STREBLIDAE_DIPTERA_HIPPOBOSCOIDEA. .

107. Breitschwerdt EB, Kordick DL. *Bartonella* infection in animals: carriership, reservoir potential, pathogenicity, and zoonotic potential for human infection. *Clin Microbiol Rev* 2000; **13**: 428–438.

108. Jiggins FM. Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics* 2003; **164**: 5–12.

109. Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci U S A* 2010; **107**: 769–774.

110. Lack JB, Nichols RD, Wilson GM, Van Den Bussche RA. Genetic signature of reproductive manipulation in the phylogeography of the bat fly, *Trichobius major*. *J Hered* 2011; **102**: 705–718.

111. Morse SF, Olival KJ, Kosoy M, Billeter S, Patterson BD, Dick CW, et al. Global distribution and genetic diversity of *Bartonella* in bat flies (Hippoboscoidea, Streblidae, Nycteribiidae). *Infect Genet Evol* 2012; **12**: 1717–1723.

112. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proc Natl Acad Sci U S A* 2014; **111**: 10257–10262.

113. Stuckey MJ, Chomel BB, de Fleurieu EC, Aguilar-Setién A, Boulouis H-J, Chang C-C. *Bartonella*, bats and bugs: A review. *Comp Immunol Microbiol Infect Dis* 2017; **55**: 20–29.

114. Gibson CM, Hunter MS. Extraordinarily widespread and fantastically complex: comparative biology of endosymbiotic bacterial and fungal mutualists of insects. *Ecol Lett* 2010; **13**: 223–234.

Supplementary Information for

Cascading effects of habitat patch area on ectoparasite communities and their associated bacterial microbiomes

SUPPLEMENTAL METHODS

DNA extraction

Flies were separated into individual tubes and washed twice by suspending in 500µL phosphate-buffered-saline (1x) and vortexing to dilute exoskeletal bacterial contamination. Following washing, the abdomen of each fly was separated from the thorax using sterile forceps and proteinase K was used to digest soft tissue from the entire fly overnight following manufacturer instructions (55°C). Extractions followed manufacturer protocol with the following exceptions: samples were bead beat in a Disruptor Genie for 20 minutes at 3000 rpm (max speed); following bead beating, samples were stored at -80°C following manufacturer guidelines; sterile water used for elution of DNA from the filter was heated to 55°C; the elution incubation step was increased to 5 minutes; and the elution step was repeated using the first eluate to re-hydrate the column filter.

16S rRNA Library Preparation: Amplification and Indexing Reactions

Earth Microbiome Project primers were employed with Illumina overhangs for barcoding Primers 515f (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGYCAGCMGCCGCGGTAA) and 806r (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACNVGGGTWTCTAAT) with Illumina overhang were used for amplification of the V4 region of 16S rRNA (Gilbert et al. 2010; Gilbert, Jansson, and Knight 2014; Apprill et al. 2015; Parada, Needham, and Fuhrman 2016). Amplicon PCRs were performed in triplicate 25µL final reaction volumes containing 10µL 5PRIME HotMasterMix (final concentrations of 1U Taq DNA Polymerase, 45mM Cl, 2.5mM Mg_2+, 200µM of each dNTP; Quantabio, Beverly, MA,USA), 5µL of each primer at 1µM concentration (final concentration of 0.2µM each), and 5µL of template DNA. Thermocycler conditions used an initial denaturation of 94°C for 2 minutes, followed by 30 cycles of 94°C for 20 seconds, 55°C for 30 seconds, and 65°C for 30 seconds, with a final elongation step at 65°C for 5 minutes and a storage temp of 4°C. Extraction negative controls were pooled into a single aliquot and an additional negative control was introduced during PCR amplification. Triplicate PCRs were combined and then cleaned using SPRIselect magnetic beads following manufacturer's instructions (Beckman Coulter, Sykesville, MD, USA). Concentration of cleaned PCR products was estimated using the Qubit 2.0 fluorometer dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA), and 10% of samples were run on a Bioanalyzer 2100 DNA High Sensitivity chip (Agilent, Santa Clara, CA, USA) to assess representative quality of preparations and verify consistency in amplicon sizes. Indexing PCRs were conducted in 50µL final reaction volumes containing 25µL KAPA HiFi HotStart Ready Mix (final concentrations of 0.5U Taq DNA polymerase, 2.5mM $MgCl_2$, and 0.3mM of each dNTP; KAPA Biosystems, Wilmington, MA, USA), 5µL of each the forward and reverse indexing primers (Illumina Nextera XT Index Kit v2, set A, set B, and set C), and 5µL clean amplicon PCR product, following recommendations in the Illumina 16S Metagenomic Sequencing Library

Preparation guidelines. Thermocycler conditions used an initial denaturation at 95C for 3 minutes, followed by 8 cycles of 95C for 30 seconds, 55C for 30 seconds, and 72C for 30 seconds, finishing with a final elongation at 72C for 5 minutes and a 4C holding temperature. Indexed libraries were cleaned using SPRIselect magnetic beads and concentration and quality of libraries was estimated as described above.

16S rRNA Library Concentration and Pooling

Libraries that were lower than 2nM were concentrated using SPRIselect magnetic beads to remove 10mM Tris pH 8.5 and vacufuged until dry. Libraries were re-hydrated with 4-6µL sterile water depending on initial concentration, and concentration and quality were re-assessed using Qubit and Bioanalyzer. Following concentration of low-yield samples, 206 equimolar libraries were combined into a 3.4nM "high concentration pool" and 23 libraries at a concentration of less than 3.4nM were combined into a single 1nM "low concentration pool", which was used to dilute the "high concentration pool" from 3.4nM to 2nM for sequencing.

Barcoding of bat flies

To barcode the bat flies for which microbiomes were sequenced, a 710bp fragment of COI was amplified from extracted DNA using universal primers developed by (Folmer et al. 1994); LCO1490: 5'-GGTCAACAAATCATAAAGATATTGG-3', HCO2198: 5'-TAAACTTCAGGGTGACCAAAAAATCA-3'). PCRs were conducted in 15µL reactions using 7.5µL 2x TopTaq Master Mix (Qiagen, Hilden, Germany); 0.1µM final concentration of each the forward and reverse primer, 1.5µL of 10x Coral Load (Qiagen, Hilden, Germany), and 1µL of template DNA, with the remaining volume filled by sdH_2O. Thermocycler conditions followed (Hebert et al. 2003). The success of PCRs was confirmed using gel electrophoresis (1.5% agarose gel). PCRs were cleaned using AMPure XP beads (Beckman Coulter, Indianapolis, IN) following manufacturer instructions. Cycle sequencing reactions were conducted in 10µL reactions containing 1µL Big Dye Terminator v3.1 (Life Technologies Corporation, Carlsbad, CA, USA), 1µL extension buffer, 0.1µM final concentration of primer, 2µL of cleaned PCR product, and sdH_20 to the final volume. Thermocycler conditions were as follows: initial denaturation at 94C for 5min, 25 cycles of denaturation at 94C for 40s, annealing at 50C for 30s, and elongation at 60C for 4min, and holding at 10C. Cycle sequencing reactions were cleaned using ethanol precipitation and rehydrated with DNA Injection Solution (Montage, Temecula, CA, USA) for sequencing on the ABI 3730xl DNA Analyzer at the Sackler Institute of Comparative Genomics at the American Museum of Natural History. Sequence chromatographs were trimmed to a 645bp segment, checked for quality, and aligned using ClustalW in Geneious v.10.2.4 (Thompson, Higgins, and Gibson 1994; Larkin et al. 2007; Kearse et al. 2012). A phylogeny was reconstructed using RAxML v.8 assuming a model of evolution of GTR+G, based on AIC scores from jModelTest 2.1, with 1000 bootstrap replicates on the CIPRES Science Gateway (Miller, Pfeiffer, and Schwartz 2010; Darriba et al. 2012; Stamatakis 2014). The phylogeny was examined in FigTree v.1.4.2 to confirm clades established from morphological identifications of bat flies (Rambaut and Drummond 2012; Rambaut 2014).

QIIME2 processing

Following de-multiplexing, samples were processed using the QIIME2 v.2018.2 pipeline (https://docs.qiime2.org/2018.2/). DADA2 was used to filter out PhiX reads and chimeras, truncate the length of reads (forward = 200bp, reverse = 180bp), and cluster reads into unique

amplicon sequence variants (ASVs) corrected for Illumina sequencing errors (Callahan et al. 2016). Reads were aligned using the MAFFT plugin in QIIME2 (FFT-NS-i;(Katoh et al. 2002, 2005; Katoh and Toh 2007). Default parameters were used to mask highly variable regions of the alignment and reconstruct a phylogeny using the FastTree2 plugin (Price, Dehal, and Arkin 2010), which was midpoint-rooted. The GreenGenes Database, v.13.5, trimmed to only the 16S rRNA V4 region, was used as a reference to train a naïve Bayes q2-feature-classifier for taxonomic identification of ASVs (DeSantis et al. 2006).

Filtering 16S rRNA data for contamination

First, any bacterial taxon detected in the negative controls was removed from all other samples, with the exception of *Arsenophonus*. This genus of bacteria contains known symbionts of insects and is expected to be associated with bat flies (Trowbridge, Dittmar, and Whiting 2006; Nováková, Hypsa, and Moran 2009). As *Arsenophonus* is highly abundant in the samples sequenced for this study, it may be that its detection in the extraction control (0.7% of 3,524 total reads) and PCR control (55% of 63 total reads) is due to index bleed, a known issue when multiplexing samples (Eisenhofer et al. 2019), and it is not treated as a contaminant here. Next, bacterial genera were removed that are known laboratory contaminants (Eisenhofer et al. 2019), as were reads that were classified as being derived from mitochondria, chloroplast, or Archaea, or those that could not be classified beyond phylum. Data were exported from QIIME2 and reformatted for import into the R package *phyloseq* v.1.26.1 (McMurdie and Holmes 2012, 2014) for further decontamination and all downstream analyses. We used the R package *decontam* to identify ASVs whose frequency is inversely correlated with initial library concentration (Davis et al. 2018). Nine additional ASVs were identified as potential contaminants and eliminated from the dataset. *Arsenophonus* was not identified as a contaminant by *decontam*.

Reference Database for "Candidatus Aschnera chinzeii"

The GreenGenes Database does not include "*Candidatus* Aschnera chinzeii", a close relative of *Arsenophonus* that has previously been identified as the primary symbiont of some nycteribiid flies (Hosokawa et al. 2012; Duron et al. 2014). To identify reads belonging to "*Candidatus* Aschnera chinzeii," we built a custom BLAST database containing reference sequences for "*Candidatus* Aschnera chinzeii" (N=4), *Arsenophonus* (N=37), and "*Candidatus* Phlomobacter" (N=3; Silva Ribosomal RNA Database; (Quast et al. 2013; Yilmaz et al. 2014; Glöckner et al. 2017) against which we compared all 16SrRNA ASVs that were classified as *Arsenophonus* by the naïve Bayes classifier.

Transformation, ordination, PERMANOVA of compositional data

Metabarcoding using high-throughput sequencing is compositional in nature – meaning the total observations (reads per ASV) for a sample contain no information about the total number of microbes and is dependent on the sequencing capacity of the instrument (Fernandes et al. 2014; Gloor and Reid 2016; Gloor et al. 2017; Tsilimigras and Fodor 2016; Xia and Sun 2017). To correct for the compositional nature of 16S rRNA sequencing data, isometric log-ratio transformations were implemented in the R package *philr* v.1.8.1 (Silverman et al. 2017). This transformation utilizes a user-provided bacterial phylogeny to standardize the abundance of bacterial taxa in a sample by the abundance of its sister taxon, creating "balances" at each node on the phylogeny (Silverman et al. 2017). Euclidean distances between *philr* balances provide

phylogenetic and abundance information about the bacteria in a sample, similar to weighted UniFrac, that can be used for ordination and down-stream statistical analysis (Gloor et al. 2017; Silverman et al. 2017).

As this is a nested system (within each fragment, we expect to see a subset of bat species, and within each bat species, only a subset of bat flies occur, and within each bat fly only a subset of bacterial taxa occur), assessing each variable separately ignores the interactions that could impact our conclusions. Sequential (Type I) sum of squares was used to account for the nonindependence of variables in testing for significant differentiation between microbiome communities. In each test, parasite species was the first variable, followed by one additional variable, and the interaction between parasite species and the additional variable (e.g., pairwise sample distance matrix ~ parasite species + log-scaled area + the interaction between parasite species and log-scaled area). The additionally examined variables were bat species, bat sex, bat individual, region (REGUA area or southern sites), log2-scaled area, log2-scaled isolation, distance from source, protection status (within REGUA or outside of REGUA, excluding the southern sites), and sampling site. PERMANOVA analyses were performed on a dataset containing all localities with taxa filtered at 0.01% relative abundance per sample, a dataset containing all localities with taxa filtered more strictly at 0.1% relative abundance per sample, only the REGUA area localities (no southern sites), and only unprotected REGUA area localities (no localities within REGUA).

Bacterial interaction network reconstruction and analysis

Southern fragments were excluded from all network analyses and REGUA area fragments F3 and F6 were excluded from the habitat fragment networks, because they had fewer than 10 samples. Species-specific networks were reconstructed for well-sampled parasite species. A within-REGUA and an outside-of-REGUA network was reconstructed for each *Aspidoptera falcata*, *Basilia juquiensis*, *Paratrichobius longicrus*, and *Strebla wiedemanni*. Only an outside-of-REGUA network was constructed for each *Strebla guajiro* and *Trichobius joblingi* because sample sizes were too low to estimate networks for these species within REGUA. The within-REGUA samples of *Strebla wiedemanni* were filtered so that only taxa that occurred at least 10 times (summed across all samples used in the network) were maintained, so that the network would reach stability.

To control for network size and shape, we created a null distribution for each habitat fragment network of 100 randomly re-wired graphs with degree distribution preserved. For each random network, the number of rewiring trials performed was equal to ten times the total number of nodes in the network. We centered the modularity of each measured network by the mean modularity of its corresponding null distribution. We also calculated the Z-score modularity using the mean and standard deviation of the measured networks (e.g., [modularity of F1 network − mean modularity of all networks]/standard deviation of modularity of all networks) and the Z-score modularity using the mean and standard deviation of each null network (e.g., [modularity of F1 network − mean modularity of F1-specific null distribution]/standard deviation of F1-specific null distribution).

LITERATURE CITED

Apprill, A., S. McNally, R. Parsons, and L. Weber. 2015. "Minor Revision to V4 Region SSU RRNA 806R Gene Primer Greatly Increases Detection of SAR11 Bacterioplankton." *Aquatic Microbial Ecology: International Journal* 75 (2): 129–37.

Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. "DADA2: High-Resolution Sample Inference from Illumina Amplicon Data." *Nature Methods* 13 (7): 581–83.

Darriba, Diego, Guillermo L. Taboada, Ramón Doallo, and David Posada. 2012. "JModelTest 2: More Models, New Heuristics and Parallel Computing." *Nature Methods* 9 (8): 772.

Davis, Nicole M., Diana M. Proctor, Susan P. Holmes, David A. Relman, and Benjamin J. Callahan. 2018. "Simple Statistical Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics Data." *Microbiome* 6 (1): 226.

DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen. 2006. "Greengenes, a Chimera-Checked 16S RRNA Gene Database and Workbench Compatible with ARB." *Applied and Environmental Microbiology* 72 (7): 5069–72.

Duron, Olivier, Ulrich E. Schneppat, Arnaud Berthomieu, Steven M. Goodman, Boris Droz, Christophe Paupy, Judicaël Obame Nkoghe, Nil Rahola, and Pablo Tortosa. 2014. "Origin, Acquisition and Diversification of Heritable Bacterial Endosymbionts in Louse Flies and Bat Flies." *Molecular Ecology* 23 (8): 2105–17.

Eisenhofer, Raphael, Jeremiah J. Minich, Clarisse Marotz, Alan Cooper, Rob Knight, and Laura S. Weyrich. 2019. "Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations." *Trends in Microbiology* 27 (2): 105–17.

Fernandes, Andrew D., Jennifer Ns Reid, Jean M. Macklaim, Thomas A. McMurrough, David R. Edgell, and Gregory B. Gloor. 2014. "Unifying the Analysis of High-Throughput Sequencing Datasets: Characterizing RNA-Seq, 16S RRNA Gene Sequencing and Selective Growth Experiments by Compositional Data Analysis." *Microbiome* 2 (May): 15.

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 (5): 294–99.

Gilbert, Jack A., Janet K. Jansson, and Rob Knight. 2014. "The Earth Microbiome Project: Successes and Aspirations." *BMC Biology* 12 (August): 69.

Gilbert, Jack A., Folker Meyer, Dion Antonopoulos, Pavan Balaji, C. Titus Brown, Christopher T. Brown, Narayan Desai, et al. 2010. "Meeting Report: The Terabase Metagenomics Workshop and the Vision of an Earth Microbiome Project." *Standards in Genomic Sciences* 3 (3): 243–48.

Glöckner, Frank Oliver, Pelin Yilmaz, Christian Quast, Jan Gerken, Alan Beccati, Andreea Ciuprina, Gerrit Bruns, et al. 2017. "25 Years of Serving the Community with Ribosomal RNA Gene Reference Databases and Tools." *Journal of Biotechnology* 261 (November): 169–76.

Gloor, Gregory B., Jean M. Macklaim, Vera Pawlowsky-Glahn, and Juan J. Egozcue. 2017. "Microbiome Datasets Are Compositional: And This Is Not Optional." *Frontiers in Microbiology* 8 (November): 2224.

Gloor, Gregory B., and Gregor Reid. 2016. “Compositional Analysis: A Valid Approach to Analyze Microbiome High-Throughput Sequencing Data.” *Canadian Journal of Microbiology* 62 (8): 692–703.

Hebert, Paul D. N., Alina Cywinska, Shelley L. Ball, and Jeremy R. deWaard. 2003. “Biological Identifications through DNA Barcodes.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270 (1512): 313–21.

Hosokawa, Takahiro, Naruo Nikoh, Ryuichi Koga, Masahiko Satô, Masahiko Tanahashi, Xian-Ying Meng, and Takema Fukatsu. 2012. “Reductive Genome Evolution, Host–Symbiont Co-Speciation and Uterine Transmission of Endosymbiotic Bacteria in Bat Flies.” *The ISME Journal* 6 (3): 577–87.

Katoh, Kazutaka, Kei-Ichi Kuma, Hiroyuki Toh, and Takashi Miyata. 2005. “MAFFT Version 5: Improvement in Accuracy of Multiple Sequence Alignment.” *Nucleic Acids Research* 33 (2): 511–18.

Katoh, Kazutaka, Kazuharu Misawa, Kei-Ichi Kuma, and Takashi Miyata. 2002. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30 (14): 3059–66.

Katoh, Kazutaka, and Hiroyuki Toh. 2007. “PartTree: An Algorithm to Build an Approximate Tree from a Large Number of Unaligned Sequences.” *Bioinformatics* 23 (3): 372–74.

Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, et al. 2012. “Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data.” *Bioinformatics* 28 (12): 1647–49.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, et al. 2007. “Clustal W and Clustal X Version 2.0.” *Bioinformatics* 23 (21): 2947–48.

McMurdie, Paul J., and Susan Holmes. 2012. “Phyloseq: A Bioconductor Package for Handling and Analysis of High-Throughput Phylogenetic Sequence Data.” *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 235–46.

———. 2014. “Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible.” *PLoS Computational Biology* 10 (4): e1003531.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. “Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees.” In *2010 Gateway Computing Environments Workshop (GCE)*, 1–8. ieeexplore.ieee.org.

Nováková, Eva, Václav Hypsa, and Nancy A. Moran. 2009. “Arsenophonus, an Emerging Clade of Intracellular Symbionts with a Broad Host Distribution.” *BMC Microbiology* 9 (July): 143.

Parada, Alma E., David M. Needham, and Jed A. Fuhrman. 2016. “Every Base Matters: Assessing Small Subunit RRNA Primers for Marine Microbiomes with Mock Communities, Time Series and Global Field Samples.” *Environmental Microbiology* 18 (5): 1403–14.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. “FastTree 2--Approximately Maximum-Likelihood Trees for Large Alignments.” *PloS One* 5 (3): e9490.

Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. “The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools.” *Nucleic Acids Research* 41 (Database issue): D590-6.

Rambaut, A. 2014. “FigTree 1.4. 2 Software.” *Institute of Evolutionary Biology, Univ. Edinburgh*.

Rambaut, A., and A. J. Drummond. 2012. “FigTree Software.”

Silverman, Justin D., Alex D. Washburne, Sayan Mukherjee, and Lawrence A. David. 2017. “A Phylogenetic Transform Enhances Analysis of Compositional Microbiota Data.” *ELife* 6 (February). https://doi.org/10.7554/eLife.21887.

Stamatakis, Alexandros. 2014. “RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies.” *Bioinformatics* 30 (9): 1312–13.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. “CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice.” *Nucleic Acids Research* 22 (22): 4673–80.

Trowbridge, Richard E., Katharina Dittmar, and Michael F. Whiting. 2006. “Identification and Phylogenetic Analysis of Arsenophonus- and Photorhabdus-Type Bacteria from Adult Hippoboscidae and Streblidae (Hippoboscoidea).” *Journal of Invertebrate Pathology* 91 (1): 64–68.

Tsilimigras, Matthew C. B., and Anthony A. Fodor. 2016. “Compositional Data Analysis of the Microbiome: Fundamentals, Tools, and Challenges.” *Annals of Epidemiology* 26 (5): 330–35.

Xia, Yinglin, and Jun Sun. 2017. “Hypothesis Testing and Statistical Analysis of Microbiome.” *Genes & Diseases* 4 (3): 138–48.

Yilmaz, Pelin, Laura Wegener Parfrey, Pablo Yarza, Jan Gerken, Elmar Pruesse, Christian Quast, Timmy Schweer, Jörg Peplies, Wolfgang Ludwig, and Frank Oliver Glöckner. 2014. “The SILVA and ‘All-Species Living Tree Project (LTP)’ Taxonomic Frameworks.” *Nucleic Acids Research* 42 (Database issue): D643-8.

Table S1: Patch area, isolation, and distance from source (REGUA).

Fragment ID	Area	Isolation	Distance From Source
F1	21.15	600	3748.024
F2	34.11	234.31	5453.124
F3	41.04	84.85	3142.76
F4	52.11	362.49	6557.559
F5	84.33	150	8520.907
F6	92.34	210	2569.115
F7	99.99	349.86	6753.698
F8	117.27	134.16	357.26
F9	184.77	174.93	7525.56
F10	228.78	480	405.114
REGUA	62378.64	60	NA
REGUA2	62378.64	60	NA
REGUA3	62378.64	60	NA

Table S2: PERMANOVA on well-sampled species showing the p-value (top; *=p<0.05, **=p<0.005, ***=p<0.0005), R2 (middle) and p-value for homoscedasticity (bottom, significance indicates violation of the assumptions of PERMANOVA).

	Paratrichobius longicrus	*Speiseria ambigua*	*Strebla guajiro*	*Trichobius joblingi*
	0.8050	0.0820	0.8723	0.0091**
Parasite Sex	0.0164	0.2128	0.0254	0.2513
	0.5820	0.0070**	0.9680	0.0110*
	0.7014	0.8746	0.0573	0.5582
Bat Sex	0.0249	0.0128	0.3233	0.0206
	0.9070	0.0790	0.0001***	0.5070
	0.6494		0.5742	0.6700
Protection Status	0.0133		0.0314	0.0073
	0.8320		0.5490	0.8750
	0.6373	0.4681	0.6516	0.5989
Log_2 Area	0.0139	0.0391	0.0248	0.0145
	0.5110	0.0630	0.4260	0.6850
	0.7385	0.7440	0.1665	0.6143
Log_2 Isolation	0.0106	0.0124	0.1042	0.0120
	0.4870	0.0580	0.4180	0.7210
	0.2919	0.9885	0.8427	0.1096
Sampling Site	0.2836	0.2705	0.2924	0.5395
	0.4750	0.0710	0.4930	0.6240

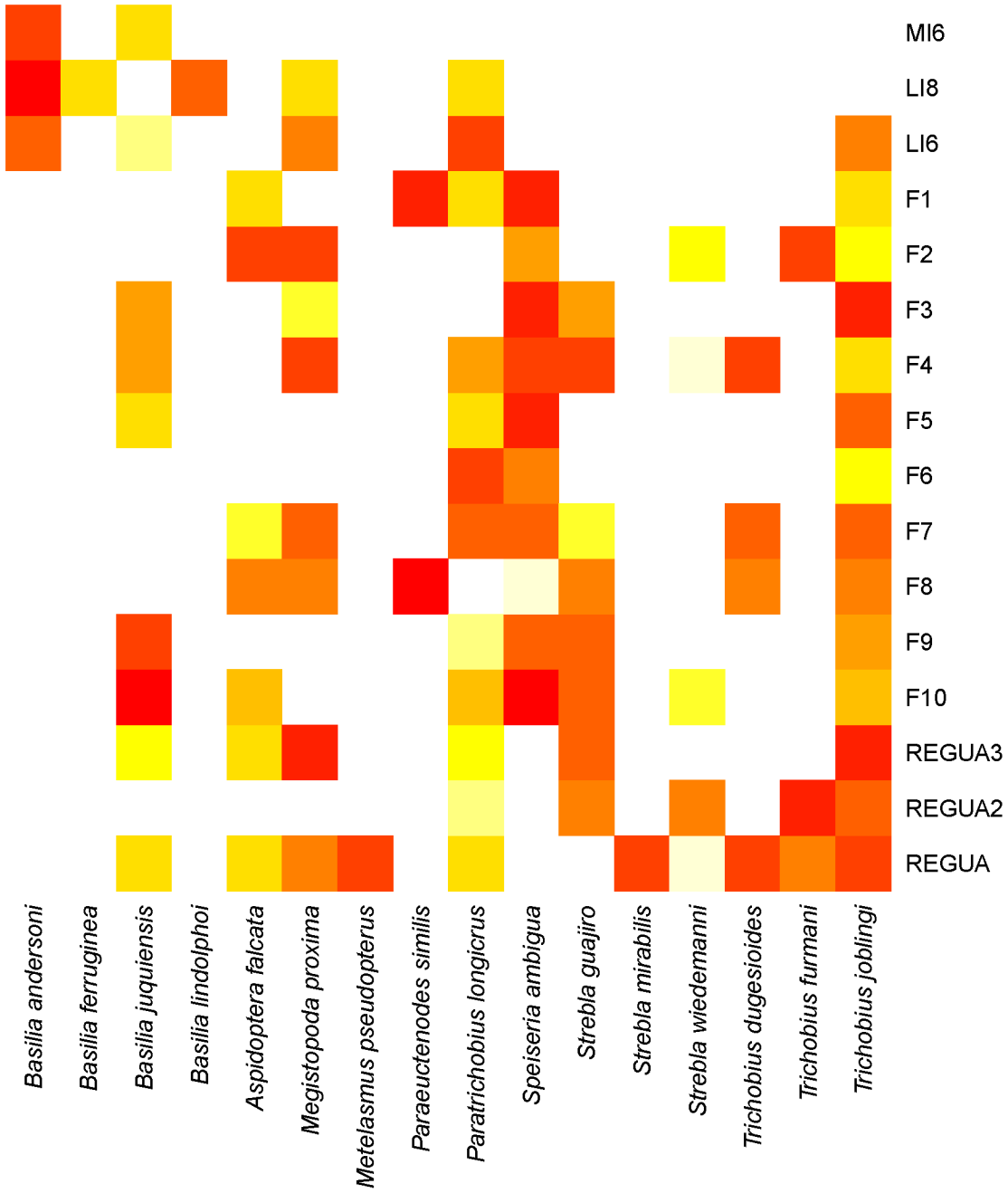


Figure S1: Heatmap shows distribution of parasite species across the sampled sites, where white means no samples were collected, red means ~1 parasite individual was collected, and pale yellow means ~5 parasite individuals were collected.

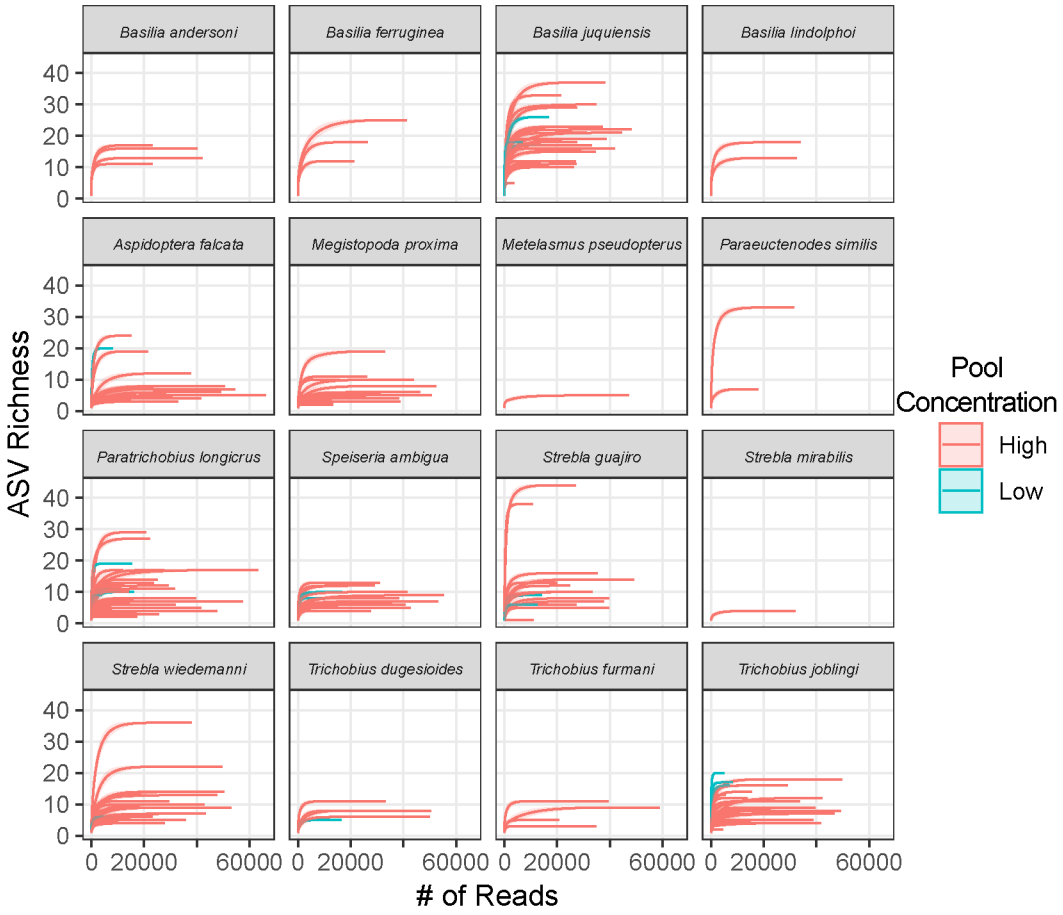


Figure S2: Rarefaction curves of bacterial ASVs detected at various sequencing depths in each parasite species. Each line represents a sample. Red lines are high concentration samples and blue lines are low concentration samples. Reads were removed from a sample if they were present at less than 0.01% relative abundance.

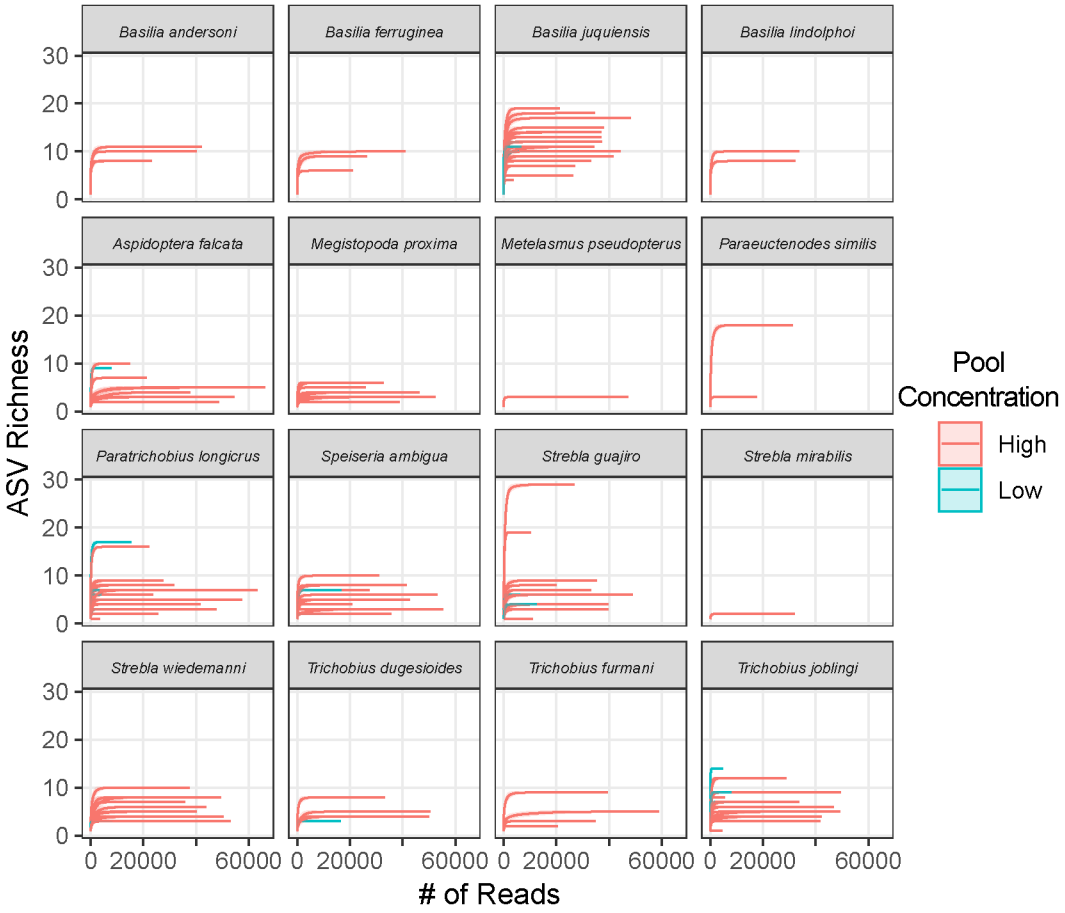


Figure S3: Rarefaction curves of bacterial ASVs detected at various sequencing depths in each parasite species. Each line represents a sample. Red lines are high concentration samples and blue lines are low concentration samples. Reads were removed from a sample if they were present at less than 0.1% relative abundance.

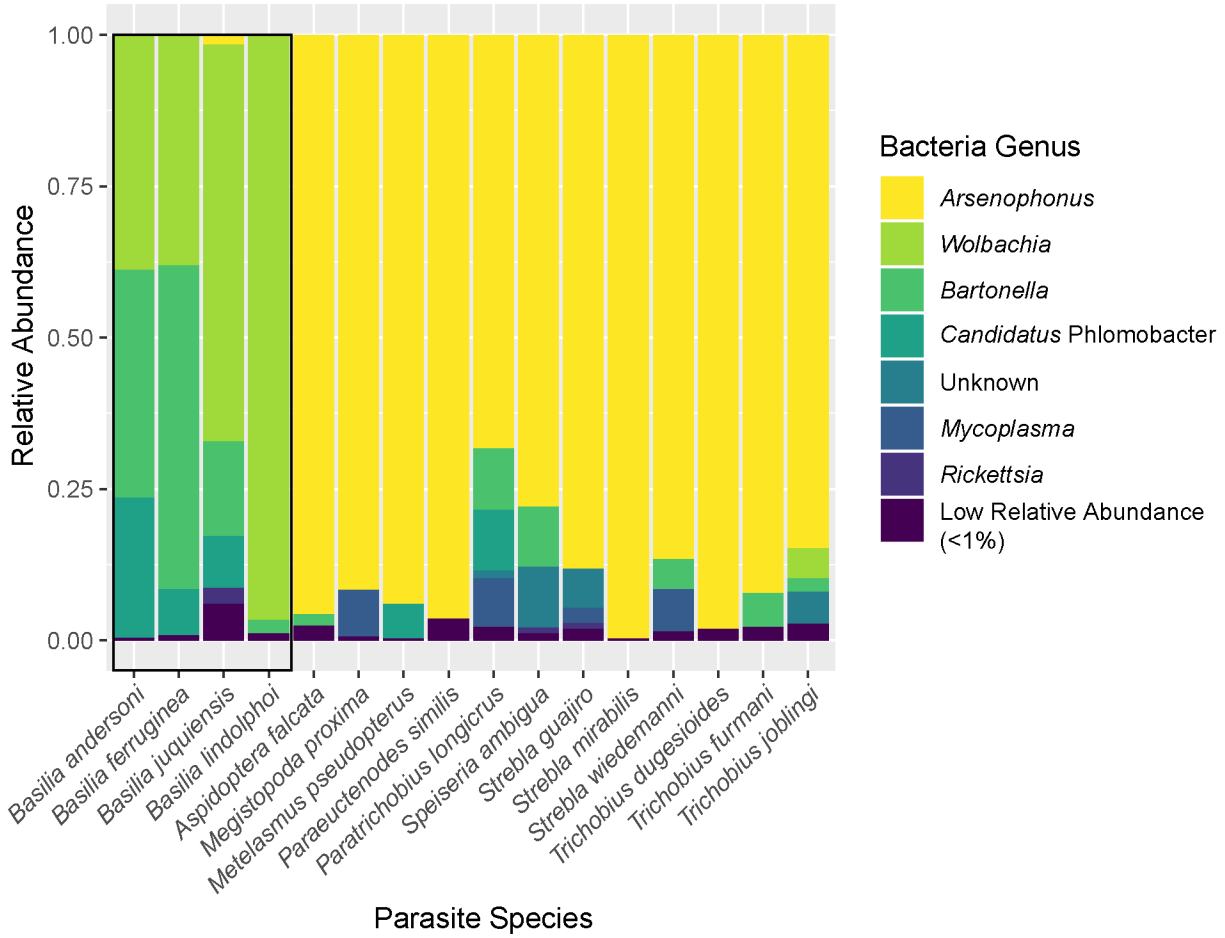
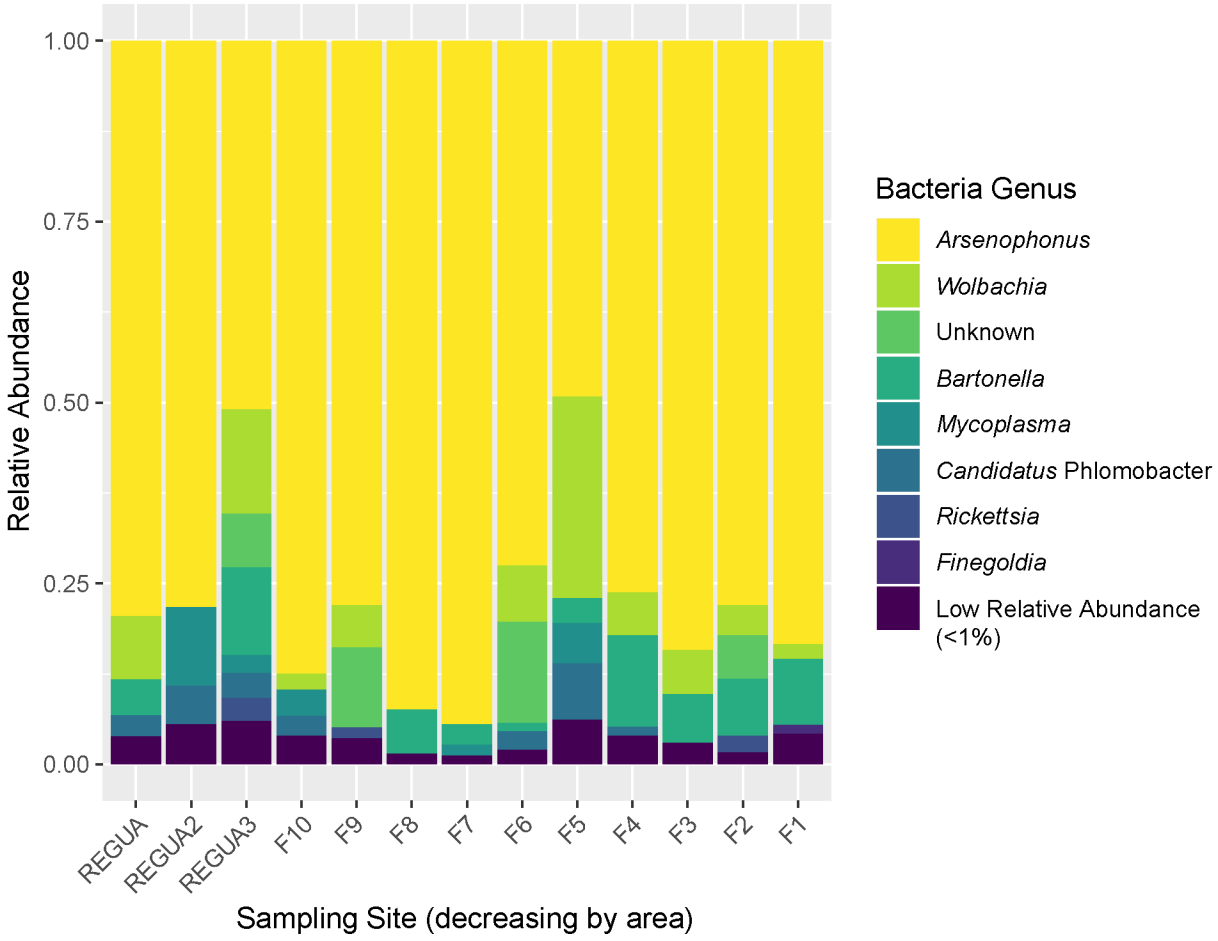
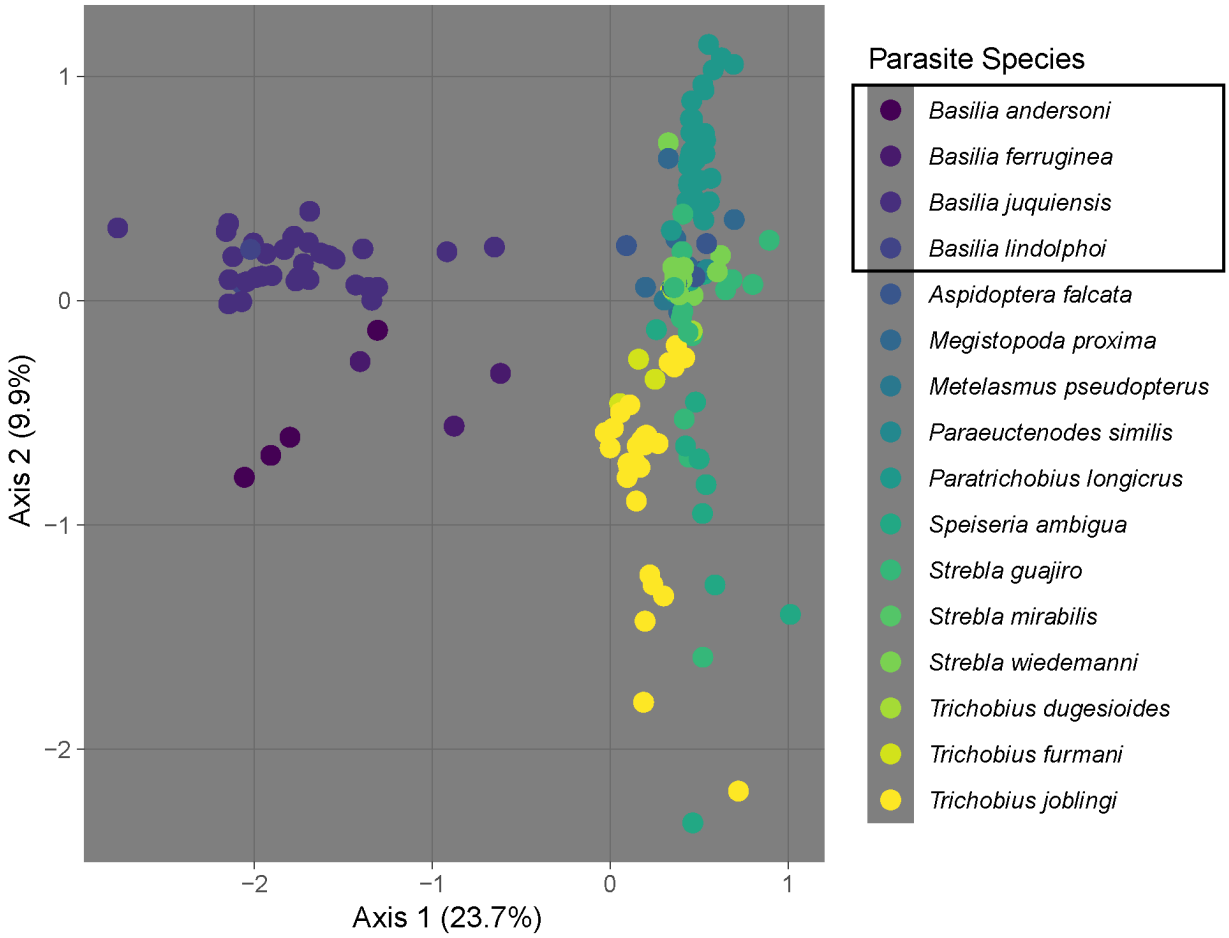


Figure S4: Relative abundance of bacterial genera in each parasite species, when reads were removed from a sample if they were present at less than 0.1% relative abundance. Nycteribiidae are within the black box and Streblidae are outside of the black box.



Figures S5: Relative abundance of bacterial genera at each sampling site.



Figures S6: Principal coordinates analysis of samples when were removed from a sample if they were present at less than 0.1% relative abundance. Colors indicate parasite species.

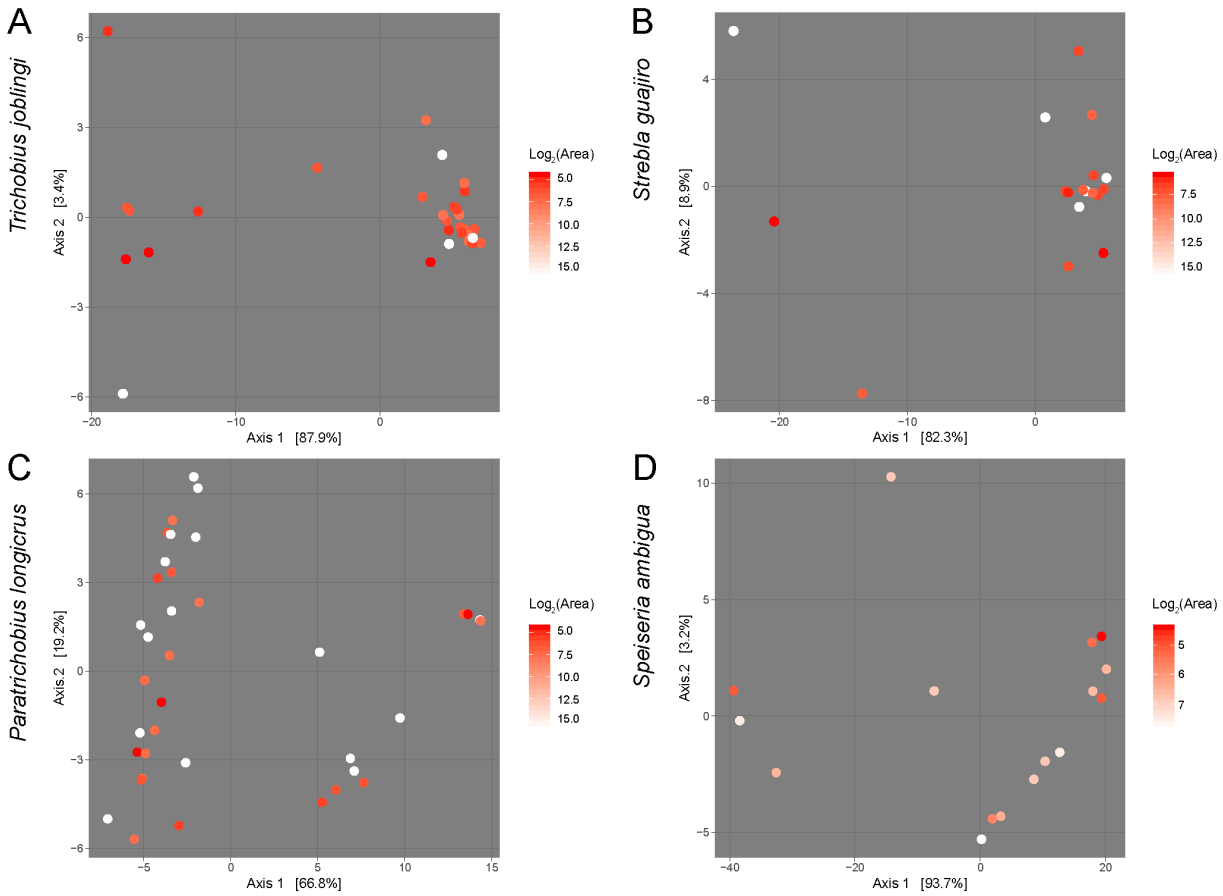


Figure S7: Principal Coordinates Analysis on the Euclidean distances between philr-transformed microbial abundances of each of the four most well-sampled bat fly species (A-D). Each species is plotted with distance to source colored from white (near) to red (far) and separately with habitat fragment area colored white (large) to red (small).

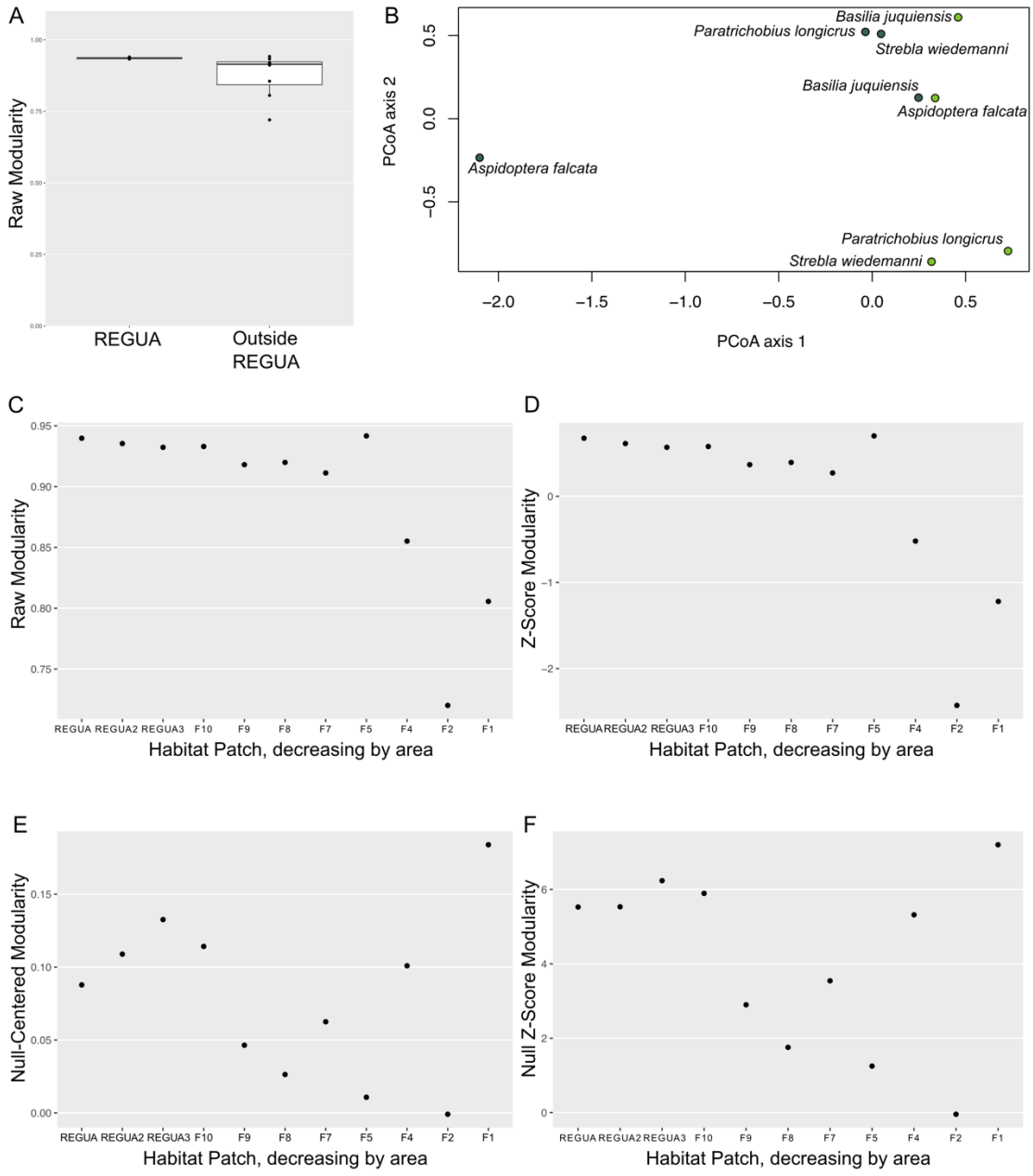


Figure S8: (A) Raw modularity between networks from within and outside of REGUA. (B) PCoA of orbit distributions of within-species networks. Lime green dots indicate networks outside of REGUA and dark green indicated within-REGUA networks. Parasite species names are provided next to each point. (C) Raw modularity of habitat patch networks decreasing by area. (D) Z-score modularity of habitat patch networks decreasing by area. (E) Null-centered modularity of habitat patch networks decreasing by area. (F) Modularity of habitat patch networks centered and scaled by the their null distributions.