MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 2617–2627 doi: 10.1111/j.1365-294X.2012.05568.x

Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography

CALEB D. PHILLIPS*, GEORGINA PHELAN*†, SCOT E. DOWD*‡, MOLLY M. McDONOUGH*, ADAM W. FERGUSON*, J. DELTON HANSON‡, LIZETTE SILES*, NICTÉ ORDÓÑEZ-GARZA*, MICHAEL SAN FRANCISCO* and ROBERT J. BAKER*

**Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA, †Manchester Metropolitan University, Manchester M15 6BH, UK, ‡Research and Testing Laboratories, Lubbock, TX 79416, USA*

Abstract

Metagenomic methods provide an experimental approach to inform the relationships between hosts and their microbial inhabitants. Previous studies have provided the conceptual realization that microbiomes are dynamic among hosts and the intimacy of relation between micro- and macroorganisms. Here, we present an intestinal microflora community analysis for members of the order Chiroptera and investigate the relative influence of variables in shaping observed microbiome relationships. The variables ranged from those considered to have ancient and long-term influences (host phylogeny and life history) to the relatively transient variable of host reproductive condition. In addition, collection locality data, representing the geographic variable, were included in analyses. Results indicate a complex influence of variables in shaping sample relationships in which signal for host phylogeny is recovered at broad taxonomic levels (family), whereas intrafamilial analyses disclosed various degrees of resolution for the remaining variables. Although cumulative probabilities of assignment indicated both reproductive condition and geography influenced relationships, comparison of ecological measures among groups revealed statistical differences between most variable classifications. For example, ranked ecological diversity was associated with host phylogeny (deeper coalescences among families were associated with more microfloral diversity), dietary strategy (herbivory generally retained higher diversity than carnivory) and reproductive condition (reproductively active females displayed more diverse microflora than nonreproductive conditions). Overall, the results of this study describe a complex process shaping microflora communities of wildlife species as well as provide avenues for future research that will further inform the nature of symbiosis between microflora communities and hosts.

Keywords: dietary evolution, host phylogeny, host reproductive microbiota, phyllostomidae

Received 20 September 2011; revision received 17 February 2012; accepted 22 February 2012

Introduction

Metagenomic studies aim to understand patterns of microbial diversity within and among various environments. Of particular interest are investigations of microflora community structure among hosts. One of the underlying reasons for this interest is the hypothesis that observed relationships between hosts and their microbial communities are the result of factors that promote symbiotic relationships via natural selection acting on both hosts and microbial communities (Ballal *et al.* 2011). The factors (explanatory variables) that are thought to shape observed structure can be considered as temporally stratified, with the evolutionary time represented in the host sample being the broadest measurable variable, whereas lifespan processes of an individual host (physiological changes associated with growth and development, health status, etc.) represent the most recent perceivable temporal aspect of the data.

Correspondence: Caleb D. Phillips, Fax: 806-742-2963;
E-mail: caleb.phillips@ttu.edu

Previous studies have documented relationships between microflora community structure and several variables. For example, Ley *et al.* (2008) documented in their study of 13 orders of mammals that host diet primarily determined microflora relationships between hosts. Similar results have been found by Muegge *et al.* (2011). By comparison, Orchman *et al.* (2010) focusing on less evolutionary time (and less dietary diversification) by sampling within the hominids found host microflora relationships to be entirely congruent with the known host phylogeny. Other studies focusing on more temporally confined variables have also found influences on the microflora including, for example, effects of living in polluted environments (Coolon *et al.* 2010), pregnancy (Collado *et al.* 2008), the effect of mother's microbiota on the nursing offspring's microflora development (Abecia *et al.* 2007), among many others.

In the current study, we report findings of a bacterial community analysis based on a 16S amplicon sequencing survey including seven families of the order Chiroptera. The species sampled from six of the families included are strictly insectivorous. The seventh family, Phyllostomidae, contains species that have diversified into several dietary strategies including insectivory, sanguivory, nectarivory, omnivory and frugivory (this diversity has arisen since the Oligocene; Baker *et al.* forthcoming; Datzmann *et al.* 2010; Rojas *et al.* 2011). In addition, during field collection, data pertaining to individual reproductive condition and collection locality were recorded. The end result was a sample of bat hosts with associated environmental variables potentially influencing host microflora communities. The specific questions that were addressed through analyses were as follows: (i) Is there a signal for host phylogeny in the microflora of bats? (ii) Can microflora community comparisons differentiate between the various feeding strategies observed within the family Phyllostomidae? (iii) Does reproductive condition have a measurable influence on microflora communities in bats similar to that postulated for humans? and (iv) Is there a geographic component to community relationships over the spatial scale considered by the data?

Materials and methods

All samples included in this study except one (a *Macrotus californicus* specimen collected in Arizona) were field-collected in Guatemala (permit 772–2010) during July and August 2010. Sampling locations included the following four Guatemalan Departments: Izabal (n = 13), Petén (n = 15), Santa Rosa (n = 19) and Chiquimula (n = 12; Fig. 1). Bats were collected using mist nets and, immediately following euthanization, the distal colon of each individual was excised and placed directly into liquid nitrogen.

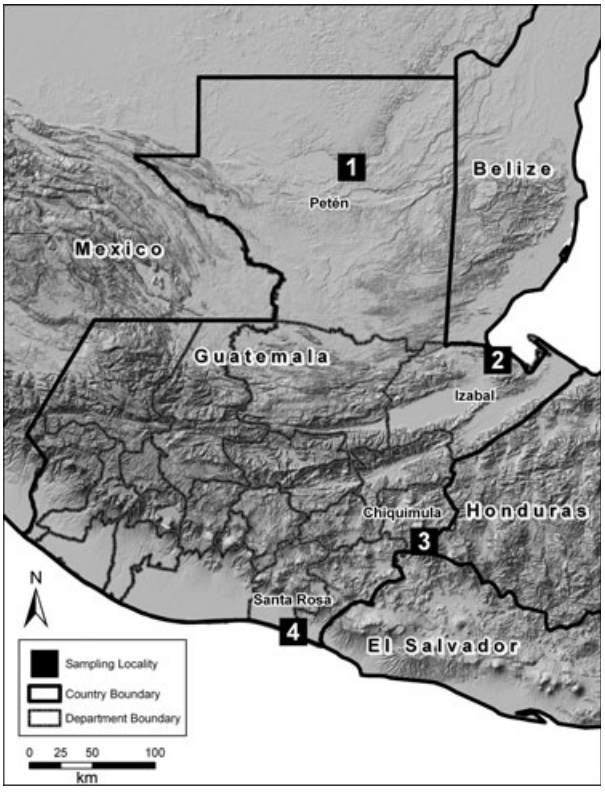

Fig. 1 Map of Guatemala in which the four sampling localities included in this study is indicated.

All individuals were classified into species in the field resulting in the identification of 27 species of bats from 23 genera. Samples were all identified as adults and included representatives of seven families of bat including the Phyllostomidae (n = 40), Emballonuridae (n = 2), Molossidae (n = 4), Mormoopidae (n = 6), Natalidae (n = 3), Noctilionidae (n = 2) and Vespertilionidae (n = 3). Among the phyllostomids, members of five dietary strategies were also represented including frugivory (n = 28), sanguivory (n = 3), insectivory (n = 2), nectarivory (n = 6) and omnivory (n = 1). In addition, within phyllostomids, all possible reproductive conditions including nonreproductive male (n = 12), nonreproductive female (n = 7), pregnant (n = 10), lactating (n = 4) and scrotal (n = 7) were represented in the sampling effort. Specific data pertaining to sample sizes and cross-classifications are summarized in Appendix I.

In the laboratory, colons were individually thawed and mucosal lining tissue of each sample was obtained by scraping with a sterile scalpel. Mucosal samples were aseptically suspended in 500 μL RLT buffer (Qiagen, Valencia, CA, USA) with β-mercaptoethanol. A sterile 5-mm steel bead (Qiagen) and 500 μL volume of

sterile 0.1-mm glass beads (Scientific Industries, Inc., NY, USA) were added to each sample to complete bacterial lyses in a Qiagen TissueLyser (Qiagen), run at 30 Hz for 5 min. Samples were then centrifuged, and 100 μL of 100% ethanol was added to a 100 μL aliquot of sample supernatant. This mixture was applied to a DNA spin column, and DNA recovery protocols were followed as instructed in the Qiagen DNA Stool Kit (Qiagen) starting at step five of the Protocol. DNA was eluted from the column in 50 μL water and diluted to a final nominal concentration of 20 ng/μL as quantified using a Nanodrop spectrophotometer (Nyxor Biotech, Paris, France).

Tag-encoded FLX amplicon pyrosequencing (bTEFAP) was performed as described previously using Gray28F 5′TTTGATCNTGGCTCAG and Gray519r 5′ GTNTTACNGCGGCKGCTG (Bailey *et al.* 2010a,b, and references therein) with primers numbered in relation to *Escherichia coli* 16S rDNA. Generation of the sequencing library was accomplished through one-step PCR with 30 cycles, and using a mixture of Hot Start and HotStar high fidelity *Taq* polymerases (Qiagen). Sequencing extended from Gray28F. bTEFAP utilized a Roche 454 FLX instrument with Titanium reagents and procedures and was performed at the Research and Testing Laboratory (Lubbock, TX, USA) based upon RTL protocols (http://www.researchandtesting.com).

Following sequencing, failed reads, low-quality ends, barcodes and primers were removed and the sequence collection was depleted of any nonbacterial ribosome sequences and chimeras using B2C2 (Gontcharova *et al.* 2010). Sequence clustering/OTU designations among the collection of sequence reads were performed with UCLUST (Edgar 2010). Clustering similarity threshold was set to 97%, minimum sequence length was 100 bp and word length (*W* value) to 20. Representative sequences for clusters were aligned to the Greengenes core set – available from http://greengenes.lbl.gov/ using PyNast (Caporaso *et al.* 2010a). For this process, the minimum percent sequence identity to include a sequence in the alignment was set to 75% and the minimum sequence length was set to 100 bp. Taxonomy was assigned using RDP classifier (Wang *et al.* 2007) with a minimum confidence to record assignment set at 0.80. Taxa summary charts were produced describing the occurrence of bacterial phylotypes within each sample.

Multiple rarefactions of the data were performed by consecutively subsampling the population of bacterial occurrences for each individual bat sample at intervals of 100 sequences between 100 and 1200. This sequential rarefaction was performed 50 times, resulting in 600 rarefied data sets. Ecological diversity was estimated by calculating the phylogenetic diversity (PD) measure described by Faith (1992). PD is a measure of the total amount of phylogenetic branch length observed within a sample, with branch length contained in a sample taken from the overall 16S rDNA gene phylogeny of all samples in the study. PD was calculated for each sample in each rarefied data set, resulting in 600 pseudoindependent PD calculations for each bat sample. PD, as opposed to one of the several other available ecological measures, was the employed statistic because the working hypothesis is that host–microfloral relationships are the result of a complex co-evolutionary/colonization phenomenon. Related, because distant branches of the microbial phylogeny on average have more dissimilar community functions (functional diversity) than closely related bacterial lineages, PD could be more powerful for understanding differences in communities when compared to alternative ecological statistics that focus on frequency and occurrence.

A UniFrac analysis was performed as originally described by Lozupone & Knight (2005). This statistic was calculated to describe the dissimilarity among bat samples by taking into account both the evolutionary distances and the frequency of occurrences of bacterial phylotypes observed among samples. For this analysis, the representative sequence set alignment was used (with gaps masked) to construct a bacterial phylogeny using FastTree (Price *et al.* 2009). This bacterial phylogeny and observation frequencies among bat samples served as input for the UniFrac analysis, resulting in a pairwise distance matrix of bat samples. These distances were displayed as either PCoAs or cladograms. Iterations of the UniFrac analysis performed using the rarefied data sets provided jackknife support values for UniFrac inferred relationships. The analytical steps described previously (with the exception of the B2C2 procedure) were performed using Qiime (Caporaso *et al.* 2010b).

Because each bat sample could be appropriately classified by multiple classification variables including dietary strategy, reproductive condition and collection locality, reads from each bat sample were pooled into appropriate classifications and part of the informatic pipeline described earlier was repeated with these groupings. The objective of this series of analyses was to provide comparison of gross community diversity within classifications. For these analyses, only samples from the family Phyllostomidae were included. The rationale for focusing on the family Phyllostomidae was twofold; (i) host sampling was primarily concentrated within this family and (ii) removing nonphyllostomid samples removed variation due to higher-level host phylogeny. Relating to this later point, an *a priori* phylogenetic congruency test (de Vienne *et al.* 2007) among phyllostomid samples was conducted, through which

signal for host phylogeny was found to be not significant (data not shown). The end result of these analyses was PD estimates for each classification. To statistically compare PD between groups, the distributions of PD estimates for each group obtained through rarefaction were compared using Kolmogorov–Smirnov (K–S) tests for location with Bonferroni correction. The K–S test, being a test for location, compares the shapes of distributions (containing information about species evenness and richness in this context; i.e. how diversity is recovered through subsampling), rather than means of distributions as a *t*-test would (a *t*-test is not appropriate here because there is no natural sample size). For clarification, because K–S tests compare shapes of distributions, these results do not signify differences in mean PD estimates. Rather, ranked PD means and distribution comparisons should be considered together in describing the overall diversity estimate, and how this diversity is distributed ecologically, respectively (the PD statistic does not incorporate aspects of evenness and richness). Furthermore, the application of this statistic here assesses similarities in sample diversity, rather than community compositions as does the UniFrac metric. In a separate analysis, to test whether members of a given classification shared more OTUs than expected by chance, G-tests for independence were performed (log-likelihood-ratio-based tests for independence; Dunning 1993) using the phyllostomid samples, not pooled by classifications. For the aforementioned analyses involving only phyllostomid samples, the *Macrotus californicus* sample from Arizona was removed due to its disparate sampling locality, as was the single sample classified as omnivorous (*Phyllostomus discolor*).

Analysis of samples pooled by family consisted of all insectivorous samples grouped by bat family (because of the uneven family sampling and because all families except Phyllostomidae were represented by insectivorous bat species). This grouping resulted in even sampling among bat families, and largely eliminated variation due to dietary strategy that otherwise would have been introduced by including noninsectivorous phyllostomid samples. The phylogenetic relationships between bat families estimated from this analysis were compared to the phylogeny of bat families presented in Teeling *et al.* (2005). The purpose of this comparison was to assess the extent to which microflora community relationships between bat families were congruent with the well-supported evolutionary relationships of hosts. The congruency index (I_{cong}) was calculated following de Vienne *et al.* (2007) to test the probability that the two phylogenies were more similar than expected by chance. In addition, to test whether PD estimated for each family contained a component of host phylogeny, a regression analysis of PD and t_{mrca} (time to most recent common ancestors from Teeling *et al.* 2005) within represented superfamily members was also performed.

To assess the extent that signal for each classification could be detected in the data set of bat samples without *a priori* classification information, cross-validated discriminant function analyses (CV-DFA) were performed. The cross-validation was performed by classifying each observation (bat sample) using a discriminant function computed from the other observations in the data set. For these calculations, the first six components from the UniFrac analysis representing approximately 75% of the variation in the data set (the remaining variation was evenly distributed among several components) were used as discriminating variables, applied to each classification separately. Component loadings were used as discriminating variables rather than the table of bacterial phylotype occurrences due to the large number of variables in this table. The UniFrac analysis including insectivorous samples (not grouped by family) was used to discriminate families, and the phyllostomid sample only UniFrac analysis was used to test discrimination of dietary strategies, reproductive conditions and collection localities (following from the same rationale described earlier). Next, cumulative binomial probabilities [$P(X \geq n)$], where X is the random probability of correct assignment and n is the observed correct assignment rate), were calculated for the frequencies of all correct discriminations. The purpose of this calculation was to provide a direct comparison of assignment rate across variables that differed in their number of categories (i.e. the number of possible classifications influences error rate in CV-DFA). Finally, Fisher's exact tests among classification variables were computed to assess the significance of association among classification variables to further guide the interpretation of discrimination statistics among variables. Discriminant function analyses and Fisher's exact tests were performed using SAS version 9.1 (SAS Institute Inc.).

Results

After culling 454 reads for low-quality and nonbacterial 16S ribosome sequences, a total of 305 183 sequences remained with an average read length of 380 bp. The average number of reads per sample was 5331 (SE = 497; Appendix I; see Supporting information Fig. S1 for rarefaction plots). A total of 7347 OTUs were recovered at the prescribed similarity clustering threshold of 97%. From these clusters, 339 bacterial phylotypes were identified. The summary chart detailing the occurrence of bacteria taxa for each sample is available for download through Table S1 (Supporting information).

The UniFrac analysis of all samples was described by 51.5% variation distributed across the first three compo-

nents of the PCoA (Fig. 2). Relationships between samples as revealed by this plot indicate that no single classification variable obviously drove sample relationships. Similarly, G-tests for association of OTUs with a given variable were all nonsignificant (Table S2, Supporting information). The overall error rate for assignment to family among insectivorous samples was 0.83, and the cumulative binomial probability (the random probability of obtaining the observed number or more correct assignments by chance) for this error rate was found to be 0.21 (cumulative binomial probabilities will subsequently be listed in parentheses directly after CV-DFA error rates). By grouping reads from individuals into family classifications (to remove intrafamilial variation), the phylogeny based on UniFrac distances was found to be statistically congruent (I_{cong} = 1.31, P = 0.04) with the accepted evolutionary relationships between bat families (Fig. 3), although not all sister relationships between families were recovered (incorrect placements of Emballonuridae and Noctilionidae).

Phylogenetic diversity was calculated for each sample and for all classifications of the data, and PD for bat families, dietary strategies and reproductive conditions are presented in Table 1 (these groupings provide the most biologically interpretable comparisons). Comparing PD values among bat families disclosed the Embal-

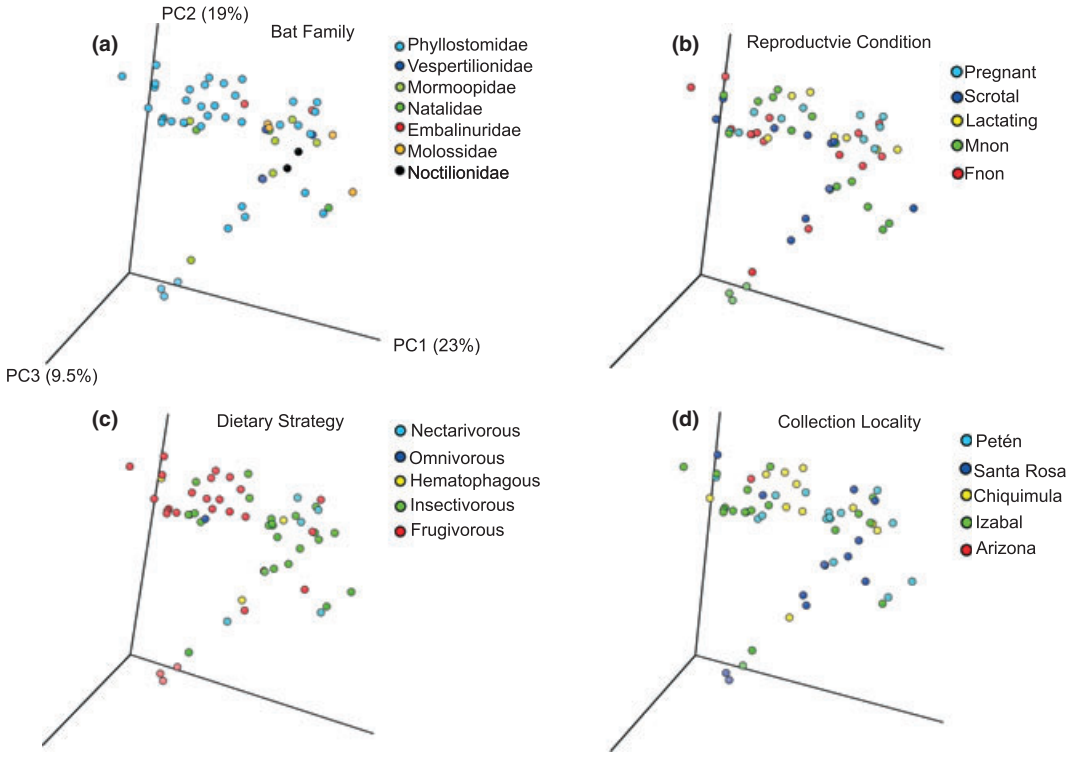

Fig. 2 Principal coordinate analysis of UniFrac metrics for all samples. Samples are colour coded by classification variables within insets a–d. For inset b, Fnon = nonreproductive female and Mnon = nonreproductive male.

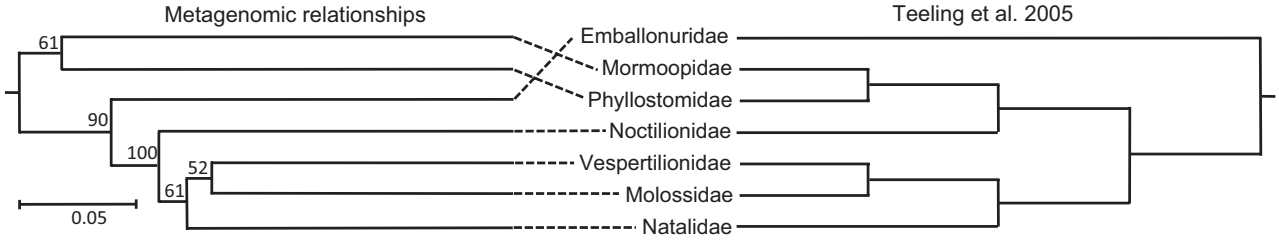

Fig. 3 Phylogeny comparison of family-level relationships inferred through microflora community relationships and the host phylogeny reported by Teeling *et al.* (2005); the two phylogenies were found to be statistically congruent (I_{cong} = 1.31, P = 0.04). Values adjacent to nodes are Jackknife support values obtained through rarefaction.

Table 1 Average PD estimate for different classifications of the data. Estimates within variable classifications that were not significantly different from each other through K–S testing are bold and italicized. Statistics for K–S tests can be found in Supporting information Tables S3–S5

Bat Family	PD
Phyllostomidae	***11.97965***
Mormoopidae	***12.50276***
Noctilionidae	***12.55025***
Molossidae	14.84584
Natalidae	16.23181
Vespertilionidae	19.05597
Emballonuridae	21.70651
Within Phyllostomids	
Dietary Strategy	
Sanguivorous	10.60437
Insectivorous	***13.5897***
Nectarivorous	***14.0515***
Frugivorous	16.66145
Reproductive Condition	
Nonreproductive male	***12.20149***
Scrotal	***12.28109***
Nonreproductive female	***12.54109***
Pregnant	16.5081
Lactating	17.01168

lonuridae as containing the most diversity, with the Phyllostomidae containing the least diversity. Furthermore, the ranking of PD values among bat families indicated a phylogenetic component of diversity. That is, the family Emballonuridae, representative of superfamily Emballonuroidea, displayed the highest estimated PD and is basal to the other superfamilies included in the analysis (t_{mrca} to other families included is approximately 55 million years ago), whereas the three families with the next highest PD estimates (representatives of superfamily Vespertilionoidea) coalesce 50 million years ago, and the final three families with the lowest PD estimates (representatives of superfamily Noctilionoidea) share a t_{mrca} 42 million years ago (Teeling *et al.* 2005). The regression analysis to assess the linearity between PD estimated for each family and the t_{mrca} to other included con-superfamilial classifications was significant ($R^2 = 0.83$, $t = -2.67$, $P = 0.04$). Furthermore, results of K–S tests were significant among all family-wise comparisons except the comparisons between members of the superfamily Noctilionoidea (Table 1; Table S3, Supporting information).

Similar to analysis among all samples, among phyllostomid sample, *G*-tests were also not significant (Table S2, Supporting information). PD estimates among dietary strategies within phyllostomid samples increased from sanguivorous, insectivorous, nectarivorous, to frugivorous. Additionally, PD estimates among dietary classifications were significantly different from each other with the exception of the insectivorous and nectarivorous classification, which were not significantly different (Table 1; Table S4, Supporting information). Although this ranking placed dietary strategies that require feeding on animals as less diverse than those that feed primarily on plant material, it should be noted that frugivory was more heavily sampled than the other three dietary strategies. Notable from a taxonomic stand-point was the observation that the Lactobacillales were in highest relative abundance in nectarivores (22.76%) and second most abundant in the frugivores (11.65%). By comparison, the occurrence of Lactobacillales in the insectivores and sanguivores was 3.2% and 1.72%, respectively (Table S1, Supporting information).

Ranked PD among reproductive conditions within phyllostomid samples was described by male categories (nonreproductive and scrotal) having lower PD when compared to all female categories (nonreproductive, pregnant and lactating). Within sexes, reproductively active categories (scrotal, pregnant and lactating) retained higher PD than their nonreproductive alternative classifications, with lactating displaying the most diversity. The results of K–S testing described significantly different PD distributions for most reproductive conditions with the exceptions that scrotal, nonreproductive male and nonreproductive female conditions were not significantly different from each other (Table 1; Table S5, Supporting information). In addition, it was found that *Bacteroides* and *Staphylococcus* were in greatest relative abundance (0.19% and 13.61%, respectively) in pregnant females as compared to any other classification (see discussion for relevance).

Assessing the contribution of dietary strategy and reproductive condition among phyllostomid samples to bacterial community relationships was described by overall error rates of assignment of 0.85 [$P(X \geq n) = 0.97$] and 0.63 [$P(X \geq n) = 0.009$], respectively. However, Fisher's exact tests indicated these two variables were significantly associated (Table Probability 4.6×10^{-6}, $P = 0.04$). Among dietary strategy discriminations, frugivores were correctly classified 59% [$P(X \geq n) = 0.0002$] of the time. Although sampling was heavily skewed towards frugivores (70% of samples were frugivores), the samples from the other three feeding strategies were never appropriately classified. Among reproductive condition discriminations, reproductively active females (lactating or pregnant) were never misclassified into a male category (nonreproductive male or scrotal) [$P(X \geq n) = 0.0002$]. In addition, nonreproductive samples (male or female) were classified into either of the two nonreproductive categories 72% [$P(X \geq n) = 0.00003$] of the time, and albeit nonreproductive males and females were classified into the correct nonrepro-

ductive gender classification at a rate of 9% [$P(X \geq n)$ = 0.91] and 29% [$P(X \geq n)$ = 0.42], respectively. Correct classification to the scrotal category occurred during 29% [$P(X \geq n)$ = 0.42) of classifications and were classified to the correct sex at a rate of 71% [$P(X \geq n)$ = 0.01]. Although male categories were assigned to the correct sex 50% [$P(X \geq n)$ = 0.59] of the time, females were classified into one of the female classifications in 79% [$P(X \geq n)$ = 0.009] of the classifications.

The overall error rate for assignment to collection locality was 0.58 [$P(X \geq n)$ = 0.03], respectively. However, there was not a clear relationship to indicate that incorrect classifications were into nearest-neighbour localities, and a *post hoc* Mantel test for indication of an isolation-by-distance phenomenon in the community data was not significant (R^2 = 0.10, Z = 436.66, P = 0.28). In addition, Fisher' exact tests yielded nonsignificant associations with diet (Table Probability = 1.21×10^{-4}, P = 0.16) and reproductive condition (Table Probability = 9.25×10^{-7}, P = 0.33), indicating that sampling did not measurably influence the analyses of collection localities.

Discussion

The interactions between micro- and macroorganisms that are thought to drive, in part, the observed patterns of species diversity and function are almost entirely unexplored. Recently, the availability of high-throughput sequencing platforms has provided an efficient tool for conducting experiments designed to document these relationships. The study of bats in this context is particularly interesting for reasons relating to their life histories. Bats being the only group of volant mammal have uniquely high energy demands that required numerous evolutionary adaptations to achieve. In addition, bats display broad diversification in obligate dietary strategies over a relatively short evolutionary interval. The objectives of the current study were to provide initial documentation on patterns of microflora community structuring across bats and to what extent variable influences were present in the data. Apparently from these results, interaction of multiple host variables drove large interindividual variation. Pertaining to this observed interindividual variation, a few aspects of the data are worth mentioning. First, the community data presented in this study were derived from mucosal scrapes of distal colons, rather than from collected faecal samples, as has generally been the case in previous studies. Although it is arguable that the analysis of the mucosal microflora community composition is more biologically relevant than that obtained from scat material, there is no data available to diagnose the differences in the data obtained between these two methods. Additionally, all of the samples used in this study were from wild-caught animals, whereas most of the samples used in previous studies that document a large effect of diet were obtained from zoos. It is not known how this difference might also influence microflora compositions.

The effect of higher-level host phylogeny on microflora arguably represents the oldest influence in this data set (although the presence of alternative reproductive conditions is clearly more ancient) as the diversification of other variables, such as diet, occurred subsequent to host divergences. It was found that by grouping insectivorous samples into family classifications, the recovered relationships (those that were best supported among all possible phylogenies) were statistically congruent with the well-supported host phylogeny presented by Teeling *et al.* (2005). However, because *G*-tests performed among samples described a nonsignificant association of OTUs with family and because overall error rate in assignment to family was large, these combined results describe large interindividual variation and the necessity for binning individuals into family groups to recover host phylogeny in the community data. In addition, relating to signal for host phylogeny, ranked PD estimates among bat families in the light of phylogeny also described an evolutionary component to observed diversity. The significant linear relationship between PD and t_{mrca} of superfamily members represented in this study implies that the evolutionary processes leading to new host lineage formation resulted in a detectable reduction in microflora diversity that was recovered proportional to the age of the lineage. Possible evolutionary phenomenon that could produce such an effect include coadaptive selective pressure, as well as a bottleneck effect acting on microflora communities as host lineages were derived. Clearly, because the divergences separating families were followed by numerous additional divergences leading to extant lineages, it appears that the detectable influence on microflora diversity by host lineage formation is mainly limited to the most ancient host nodes in the data set. However, among the superfamily Noctilionoidea, a similar pattern is recovered in which represented families with older t_{mrca} also have higher PD estimates. A notable aspect of the data is the observation that the family with the lowest overall PD estimate, the Phyllostomidae, is also the only family in which insectivores are known to take some plant material in their diet, whereas all other families consist of strict insectivores. Additional investigation in which sampling was extended to include all bat families would provide a mechanism to test the relationship between diversity and lineage formation indicated by the findings presented in this study.

Although the cumulative probability for the rate of correct assignment to feeding strategies was the highest

among variables, the fact that sampling was heavily skewed towards frugivorous samples likely hindered the formation of appropriate discriminating functions for feeding strategies. In addition, the significant association between diet and reproductive condition also likely obscured signal for resolving components of diet. Because an *a priori* expectation of the data could be resolution for differentiating among dietary strategies (see Ley *et al.* 2008; Muegge *et al.* 2011), new data directed towards smoothing sampling across dietary strategies would improve the understanding of microflora community relationships between the recently radiated dietary strategies. Such an analysis would be a valuable contrast to previous studies in that it would inform the time frame that microflora communities respond to dietary evolution pressures. In spite of sampling limitations, biologically notable patterns pertaining to diet did emerge from analyses. For example, there was an indication that herbivorous hosts (particularly frugivores) retained more microflora diversity than carnivorous hosts, as was similarly documented by Ley *et al.* (2008) in their study among mammalian orders. Furthermore, the observation that the Lactobacillales were much more abundant in herbivores than carnivores indicates the potential that this group is providing a digestive service to herbivores, perhaps similar to that of Famularo *et al.* (2005) who postulated that lactic bacteria could improve nutrient acquisition in vegetarian diets. Although dietary and reproductive classifications were statistically associated, the table of occurrences did not indicate that the association of lactic bacteria with herbivory was confounded by reproductive condition (Table S1, Supporting information). Metatranscriptome analyses are needed to understand the basis for the observed community structure among dietary strategies and how microflora community function has evolved with host diversification events.

The observation that the cumulative binomial probability for reproductive condition correct assignment was the lowest among all variables indicated this variable's importance in shaping microflora community changes within host lifespan. This is especially evident in the classification of female reproductive categories as they were never misclassified into a male category. This result indicates a relationship between female reproductive status and microflora community structure. Moreover, it was found that reproductively active female categories displayed statistically different distributions for diversity estimates (and higher mean estimates) from all other groups, whereas the microflora diversity of scrotal males, nonreproductive males, and nonreproductive females was not significantly different from each other. Similar to the current results, the study of Collado *et al.* (2008) directing experimental design towards understanding differences between obese and normal weight pregnant human females documented that microbiota compositions changed over the course of pregnancies for both classifications. Furthermore, Collado *et al.* (2008) documented an increase in the number of bacteria (the two genera considered in this study were *Bacteroides* and *Staphylococcus*) over pregnancy in both classifications (as measured by FCM-FISH and qPCR). As a tentative explanation for the bacterial increases, these authors suggested increased microflora nutrient acquisition by improving the hydrolysis of indigestible polysaccharides. Similarly, in the current study, these two genera were found to be in greatest relative abundance in pregnant females as compared to any other classification. It is not postulated here that these two bacterial genera are the important functional microflora community distinctions between reproductive and nonreproductive females; rather, this pattern indicates the relationship between reproduction and microflora community structure could occur similarly between humans and bats. Given the increased energy demands of any reproductive individual (scrotal, pregnant or lactating) relative to nonreproductive individuals, we hypothesize that microflora community differences assort with reproductive conditions reflecting an increased need for nutrient energy. Testing this hypothesis directly will require an experimental design utilizing both microflora community data from samples focusing on this question and metatranscriptomic approaches to quantify expression of metabolically relative genes within the metagenomes of the various reproductive conditions.

Given that collection localities were separated by only 100–300 km, an unexpected outcome from this study was the observation that collection locality apparently influenced host microflora. However, the direction of incorrect CV-DFA assignments as well as a *post hoc* Mantel test for a pattern of isolation by distance provided no indication that geographically adjacent collection localities were more similar to each other than to more distant collection localities. Previous studies that have explicitly investigated the biogeography of microbial communities have been both focused towards environmental sampling, rather than microbiome, and have focused biogeographic analyses at larger spatial scales (Nemergut *et al.* 2011 and references therein), inhibiting direct comparisons between these types of studies and the current study. Assuming that an unknown factor has not falsely shaped the structure of the data and noting the nonisolation-by-distance pattern among localities, it is possible that local ecosystem components among collection localities have influenced resident hosts' microflora. To understand the degree to which microflora communities can differ over fine geographic scales (comparable to those presented in this study), future studies would need to be

designed specifically to address this question. Such an experimental design would involve controlling for the evolutionary component by confining sampling to conspecific hosts, including only samples from hosts that are identified to the same gender, reproductive condition, health status, and would require that detailed documentation be taken for each collection locality that describes the local environments.

The experimental design of this study included a sample of bat hosts encompassing aspects of host phylogeny, life history, physiology and geography. These variables were appropriately viewed in a temporal setting in which higher-order host phylogeny represented the oldest temporal aspect of the data, dietary strategy evolution a more recent phenomenon, collection locality being influenced by dispersal and niche requirements, and reproductive conditions being the result of lifespan physiological processes. Results indicated that microflora community relationships between bat hosts are a reflection of overlapping influences among these variables. Although the apparent complexity of variable interaction manifested as generally high error rates for assignment, probabilities for overall assignment described the measureable influences of variables to community structure. Specific findings of interest were the indication that female reproductive condition influences microflora community diversity and that collection locality differences were detectable over relatively short geographic distances. Finally, the observation that the phylogenetic congruence observed among bat families was associated by an ecological component in which microflora community diversity increased with coalescence time of superfamily members provided insight into the evolutionary processes shaping communities over this time frame. Experimental design of future studies was also outlined in which sampling will focus on understanding individual variables in relative isolation to others. These types of studies will greatly improve our understanding of how microflora community structure and evolution occurs in nature.

Acknowledgements

Funding for this project was provided by J. Sowell, J. Matson, S. Parsons, W. Bulmer, R. Eckerlin, Natural Sciences Research Laboratory (NSRL) at Museum of Texas Tech University, Department of Biological Sciences at Texas Tech University, and the Research and Testing Laboratories, of Lubbock Texas. Field and collecting equipment were provided by Museo de Historia Natural de la Universidad de San Carlos de Guatemala and NSRL. Field work was facilitated by housing and hospitalities of the families of E. Garza, O. Vargas, J. García, and people of Lagunitas community. Museum cururation was provided by H. Garner and K. McDonald (NSRL). Additional important support was provided by personnel of Centro de Estudios Conservacionistas at Biotopo Chocón Machacas, Biotopo El Zotz, especially F. Moya, M. F. Bracamonte, M. Barrios and O. Santos. Collecting permit 772-2010 was kindly provided by E. Flores at Consejo Nacional de Áreas Protegidas CONAP, Guatemala.

References

Abecia L, Fondevila M, Balcells J, McEwan NR (2007) The effect of lactating rabbit does on the development of the caecal microbial community in the pups they nurture. *Journal of Applied Microbiology*, **103**, 557–564.

Bailey MT, Dowd SE, Parry NM, Galley JD, Schauer DB, Lyte M (2010a) Stressor exposure disrupts commensal microbial populations in the intestines and leads to increased colonization by *Citrobacter rodentium*. *Infection Immunology*, **78**, 1509–1519.

Bailey MT, Walton JC, Dowd SE, Weil ZM, Nelson RJ (2010b) Photoperiod modulates gut bacteria composition in male Siberian hamsters (*Phodopus sungorus*). *Brain Behavior Immunity*, **24**, 577–584.

Baker RJ, Bininda-Emonds ORP, Mantilla-Meluk H, Porter CA, Van Den Bussche RA (forthcoming) Molecular timescale of diversification of feeding strategy and morphology in new world leaf-nosed bats (Phyllostomidae): a phylogenetic perspective. In: *Evolutionary history of bats: fossils, molecules and morphology* (eds Gunell and Simmons), pp. 385–409. Cambridge University Press, Cambridge.

Ballal SA, Gallini CA, Segata N, Huttenhower C, Garrett WS (2011) Host and gut microbiota symbiotic factors: lessons from inflammatory bowel disease and successful symbionts. *Cellular Microbiology*, **13**, 508–517.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, **26**, 266–267.

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

Collado MC, Isolauri E, Laitinen K, Salminen S (2008) Distinct composition of gut microbiota during pregnancy in overweight and normal-weight women. *The American Journal of Clinical Nutrition*, **88**, 894–899.

Coolon JD, Jones KL, Narayanan S, Wisely SM (2010) Microbial ecological response of the intestinal flora of *Peromyscus maniculatus* and *P. leucopus* to heavy metal contamination. *Molecular Ecology*, **19**, 67–80.

Datzmann T, von Helversen O, Mayer F (2010) Evolution of nectarivory in phyllostomid bats (Phyllostomidae Gray, 1825, Chiroptera: Mammalia). *BMC Evolutionary Biology*, **4**, 165–179.

Dunning T (1993) Accurate methods for the statistics of surprise and coincidence. *Computational Lunguistics*, **19**, 61–74.

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

Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation*, **61**, 1–10.

Famularo G, De Simone C, Pandey V, Sahu AR, Minisola G (2005) Probiotic lactobacilli: an innovative tool to correct the malabsorption syndrome of vegetarians? *Medical Hypotheses*, **65**, 1132–1135.

Gontcharova V, Youn E, Wolcott RD, Hollister EB, Gentry TJ, Dowd SE (2010) Black box chimera check (B2C2): a windows-

based software for batch depletion of chimeras from bacterial 16S rRNA gene datasets. *Open Microbiology Journal*, **4**, 6.

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

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

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

Nemergut DR, Costello EK, Hamady M *et al.* (2011) Global patterns in the biogeography of bacterial taxa. *Environmental Microbiology*, **13**, 135–144.

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

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

Rojas D, Vale A, Ferrero V, Navarro L (2011) When did plants become important to leaf-nosed bats? Diversification of feeding habits in the family Phyllostomidae. *Molecular Ecology*, **20**, 2217–2228.

Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ (2005) A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science*, **307**, 580–584.

de Vienne DM, Giraud T, Martin OC (2007) A congruence index for testing topological similarity between trees. *Bioinformatics*, **23**, 3119–3124.

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

C.D.P. studies the evolution of host-microflora relationships, evolutionary developmental biology of craniofacial morphology, and phylogeography and demography of cetaceans and pinnipeds. S.E.D. is interested in the microbial ecology of diverse environments and development of new molecular and bioinformatic methods to better elucidate the same. M.M.M. is interested in mammalian systematics and phylogenetics. A.W.F.'s research focuses on niche dynamics, ecological genetics, and phylogeography of mammalian mesocarnviores. L.S. is interested in chiropteran systematics, biogeography, and conversation. N.O.-G. is interested on the systematics and evolution of small mammals of Mesoamerica. M.S.F. studies microbe-host interactions in disease and mutualism, and virulence factor gene expression. R.J.B. is interested in understanding the genetic system that resulted in insect, nectar, fruit, and blood feeders evolving from a common ancestor in less than 15 million years.

Appendix I. Summary of samples included in this study including information on variable classifications, barcodes and the number of sequence reads obtained from each sample.

Family	TK number	Binomen	Dietary Strategy	Reproductive Condition	Guatemalan Department	Barcode	Reads
Emballonuridae	TK 169180	*Saccopteryx bilineata*	Insectivore	Nonreproductive Female	Izabal	ATACGAGT	1435
	TK 169304	*Saccopteryx bilineata*	Insectivore	Nonreproductive Male	Petén	ATCAAAGT	2548
Molossidae	TK 169256	*Molossus rufus*	Insectivore	Scrotal	Petén	AGCTGCTT	3604
	TK 169358	*Nyctinomops laticaudatus*	Insectivore	Lactating	Petén	AGCTATGT	4555
	TK 169359	*Nyctinomops laticaudatus*	Insectivore	Lactating	Petén	ATAGAGGT	730
	TK 169369	*Molossus rufus*	Insectivore	Scrotal	Petén	ATATGGCT	3897
Mormoopidae	TK 169112	*Pteronotus parnellii*	Insectivore	Nonreproductive Female	Chiquimula	ACTGCCAT	496
	TK 169146	*Pteronotus parnellii*	Insectivore	Nonreproductive Female	Chiquimula	AGCTCTCT	2105
	TK 169174	*Pteronotus parnellii*	Insectivore	Nonreproductive Female	Izabal	AGGATAGT	4689
	TK 169175	*Pteronotus parnellii*	Insectivore	Nonreproductive Female	Izabal	ATAGTAGT	8631
	TK 169208	*Pteronotus davyi*	Insectivore	Nonreproductive Female	Izabal	ATAGGTCT	5163
	TK 169248	*Mormoops megalophylla*	Insectivore	Nonreproductive Male	Petén	ACTGACCT	1619
Natalidae	TK 169301	*Natalus mexicanus*	Insectivore	Nonreproductive Female	Petén	ACTGGATT	3908
	TK 169302	*Natalus mexicanus*	Insectivore	Nonreproductive Male	Petén	ATATTGAT	11916
	TK 169303	*Natalus mexicanus*	Insectivore	Nonreproductive Male	Petén	AGCTTCGT	4984
Noctilionidae	TK 169395	*Noctilio albiventris*	Insectivore	Nonreproductive Male	Santa Rosa	ACTCTGTT	4201
	TK 169396	*Noctilio albiventris*	Insectivore	Nonreproductive Female	Santa Rosa	ATATTATT	5323
Phyllostomidae	TK 163825	*Macrotus californicus*	Insectivore	Nonreproductive Male	Arizona, USA	AGGACCAT	3029
	TK 169022	*Glossophaga commissarisi*	Nectarivore	Scrotal	Chiquimula	ATACTACT	4300
	TK 169023	*Carollia sowelli*	Frugivore	Scrotal	Chiquimula	ATAGACCT	7050
	TK 169028	*Carollia sowelli*	Frugivore	Lactating	Chiquimula	AGCTGTAT	2536
	TK 169029	*Sturnira ludovici*	Frugivore	Nonreproductive Male	Chiquimula	ATCACACT	2954
	TK 169114	*Carollia sowelli*	Frugivore	Nonreproductive Male	Chiquimula	ATAGCTGT	2671
	TK 169115	*Carollia sowelli*	Frugivore	Nonreproductive Female	Chiquimula	AGGATTAT	5789
	TK 169116	*Carollia sowelli*	Frugivore	Lactating	Chiquimula	AGGCAGTT	2617

Appendix 1 : *Continued*

Family	TK number	Binomen	Dietary Strategy	Reproductive Condition	Guatemalan Department	Barcode	Reads
TK 169147	*Chiroderma salvini*	Frugivore	Pregnant	Chiquimula	ATCAACTT	5427	
TK 169148	*Artibeus jamaicensis*	Frugivore	Scrotal	Chiquimula	AGGACTGT	10015	
TK 169171	*Dermanura glauca*	Frugivore	Scrotal	Izabal	ATAGTTAT	5835	
TK 169173	*Lophostoma evotis*	Insectivore	Nonreproductive Female	Izabal	AGCTTGTT	9909	
TK 169181	*Chiroderma villosum*	Frugivore	Pregnant	Izabal	AGCTAGCT	9744	
TK 169182	*Artibeus lituratus*	Frugivore	Nonreproductive Male	Izabal	ATAGCGCT	14118	
TK 169183	*Dermanura watsoni*	Frugivore	Nonreproductive Male	Izabal	AGGCCAAT	4766	
TK 169207	*Artibeus jamaicensis*	Frugivore	Nonreproductive Male	Izabal	AGCTACAT	3683	
TK 169209	*Glossophaga commissarisi*	Nectarivore	Nonreproductive Male	Izabal	AGGCAACT	4685	
TK 169211	*Chiroderma villosum*	Frugivore	Pregnant	Izabal	AGGAGGAT	9992	
TK 169212	*Artibeus lituratus*	Frugivore	Nonreproductive Male	Izabal	AGGAGTCT	19400	
TK 169213	*Artibeus intermedius*	Frugivore	Pregnant	Izabal	ATACGTAT	2894	
TK 169214	*Artibeus jamaicensis*	Frugivore	Nonreproductive Female	Izabal	ATATGTGT	5590	
TK 169215	*Artibeus lituratus*	Frugivore	Nonreproductive Male	Izabal	ATACTCGT	7004	
TK 169216	*Uroderma bilobatum*	Frugivore	Pregnant	Izabal	AGCTCATT	4572	
TK 169217	*Uroderma bilobatum*	Frugivore	Nonreproductive Female	Izabal	AGCTTACT	7281	
TK 169250	*Phyllostomus discolor*	Omnivore	Lactating	Petén	ATATCGGT	3275	
TK 169257	*Centurio senex*	Frugivore	Pregnant	Petén	ATAGGATT	2081	
TK 169258	*Centurio senex*	Frugivore	Pregnant	Petén	AGCTGAGT	7484	
TK 169259	*Desmodus rotundus*	Sanguivore	Scrotal	Petén	AGGCACGT	11497	
TK 169346	*Desmodus rotundus*	Sanguivore	Lactating	Petén	ACTCTCGT	937	
TK 169375	*Glossophaga soricina*	Nectarivore	Pregnant	Santa Rosa	ATACTGTT	1783	
TK 169376	*Glossophaga soricina*	Nectarivore	Pregnant	Santa Rosa	ATAGTCTT	1099	
TK 169380	*Uroderma bilobatum*	Frugivore	Nonreproductive Female	Santa Rosa	AGGATCTT	8437	
TK 169381	*Uroderma bilobatum*	Frugivore	Nonreproductive Female	Santa Rosa	ATCAATAT	2561	
TK 169403	*Desmodus rotundus*	Sanguivore	Scrotal	Santa Rosa	ACTGAAAT	2541	
TK 169414	*Glossophaga soricina*	Nectarivore	Nonreproductive Male	Santa Rosa	ACTGATTT	2093	
TK 169415	*Artibeus jamaicensis*	Frugivore	Nonreproductive Male	Santa Rosa	AGGAGATT	11324	
TK 169416	*Carollia perspicillata*	Frugivore	Nonreproductive Male	Santa Rosa	ATACGCTT	8052	
TK 169417	*Artibeus lituratus*	Frugivore	Nonreproductive Female	Santa Rosa	AGCTCGAT	11907	
TK 169418	*Uroderma bilobatum*	Frugivore	Scrotal	Santa Rosa	ACTGCTGT	2209	
TK169210	*Glossophaga commissarisi*	Nectarivore	Pregnant	Izabal	ATAGCCAT	352	

Family	TK number	Genus	Dietary Strategy	Reproductive Condition	Guatemalan Department		Reads
Vespertilionidae	TK 169111	*Myotis* cf. *keaysi*	Insectivore	Lactating	Chiquimula	ACTGAGGT	10779
	TK 169338	*Eptesicus furinalis*	Insectivore	Scrotal	Petén	ATATGCAT	3043
	TK 169397	*Myotis* sp.	Insectivore	Nonreproductive Male	Santa Rosa	ACTGCGCT	2759

Data accessibility

Sequence data are available through GenBank Short Read Archive accession number SRA050160.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Rarefaction plots for each sample in which PD is plotted against number of reads (subsampled between 100 and 1200 at intervals of 100).

Table S1 Relative frequencies for all bacteria phylogroups by host sample that also includes rows of host classifications.

Table S2 Results of G testing for significant association of bacterial OTUs with host variables.

Table S3 Results of K–S testing for differences in location of distributions of PD for families obtained through rarefaction.

Table S4 Results of K–S testing for differences in location of distributions of PD for dietary strategies obtained through rarefaction.

Table S5 Results of K–S testing for differences in location of distributions of PD for reproductive conditions obtained through rarefaction.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.