Integrative and Comparative Biology

Integrative and Comparative Biology, pp. 1–13
doi:10.1093/icb/icx011

Society for Integrative and Comparative Biology

SYMPOSIUM

Microbiome Structural and Functional Interactions across Host Dietary Niche Space

Caleb D. Phillips,[1,*,†] John Hanson,[‡] Jeremy E. Wilkinson,[‡] Lawrence Koenig,[‡] Eric Rees,[‡] Paul Webala[§] and Tigga Kingston[*]

*Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA; [†]Natural Science Research Laboratory, The Museum, Texas Tech University, Lubbock, TX 79409, USA; [‡]Research and Testing Laboratory, RTL Genomics, Lubbock, TX 79407, USA; [§]Department of Tourism and Wildlife, Maasai Mara University, 20500, Narok, Kenya

From the symposium "With a Little Help from My Friends: Microbial Partners in Integrative and Comparative Biology (SICB wide)" presented at the annual meeting of the Society for Integrative and Comparative Biology, January 4–8, 2017 at New Orleans, Louisiana.

[1]E-mail: caleb.phillips@ttu.edu

Synopsis Host-associated microbiomes are integral components of host health, but microbiome community structure varies among and within hosts. Reconciling community variability with the apparent dependence of hosts on community function, and characterizing how functional divergence proceeds across niches, remains challenging. Here, through the study of gut microbiomes and diets of three insectivorous bat species we characterize how community structure is shaped by predicted functional properties of community members. We found that while host diet and microbiome community composition do not significantly relate to each other, host diet and metagenome function do, suggesting that diet directly selects metagenomic functions rather than communities. We use a novel inference framework to show how the discordance between community structure and functional variation derives from functional equivalence and is influenced by the continuum of shared and derived gene sets across microbial lineages. Our findings help clarify how metagenome community structure–function relationships contribute to deterministic processes in community assembly, and describe the basis for metagenomic differences across ecologically similar hosts.

Introduction

Microbiomes, the microbial communities inhabiting environments, are nearly ubiquitous on Earth, inhabiting soil, water, ice, and extreme environments, as well as a variety of external and internal surfaces of macro-organisms (Lozupone and Knight 2005). Host-associated microbiomes are often considered extensions of their hosts, reflecting the contribution of microbial communities to host tissue structural integrity (Kumar and Mason 2015), wound healing (Wolcott et al. 2016), immune function (Round and Mazmanian 2009; Lathrop et al. 2011), and nutrient acquisition (Sommer et al. 2016), among others. Generally, the maintenance of these communities at host-optimal compositions is associated with health and fitness of hosts (e.g., Turnbaugh et al. 2007), while disruption of community structure (i.e., dysbiosis) has commonly been shown to negatively affect host health and limit microbiome services (e.g., Turnbaugh and Gordon 2009).

In terms of biomass, gut microbiomes are the most abundant microbial systems co-occurring with mammals, and are largely a consequence of both the nutrient-rich environment of mammalian digestive systems and the beneficial functions these communities provide to their hosts (Muegge et al. 2011). Primarily, digestive microbiomes benefit hosts by providing nutrients through both catabolic and anabolic pathways (Hollister et al. 2014). Trends in community composition across digestive microbiomes of mammalian species are in large part explained by the host's dietary guild (Ley et al. 2008; Phillips et al. 2012); there are consistent differences in gut-microbiome diversity among dietary strategies (e.g., carnivory vs. herbivory). However, differences in microbial metagenome structure and function

among host species that occupy overlapping dietary niche space is less clear (Bolnick et al. 2014). This is an important consideration because the ability of microbiome communities to respond to modest trait differences among hosts should facilitate host adaptive success. Given the large size of metagenomes relative to genomes, sufficient metagenomic variation should exist to allow for the formation of locally optimized microbial communities. Yet, how the host environment, compositional variability, and functional repertoire interact is poorly understood.

The composition of ecological communities is governed by the interplay of filtered colonization of species from a historically-constrained regional pool, interactions among species within the community, and dispersal among communities (Leibold et al. 2004). These processes may be deterministic, that is influenced by species attributes, or stochastic and hence neutral with respect to species' traits (Hubbell 2006). More specifically, niche-appropriation theory posits that observed community structure is the result of competition among potential community members for niche space in a given environment (Ricklefs and Travis 1980). Trait combinations that confer a competitive advantage to species are retained in the community, and are ultimately the basis for selection (Schmidt et al. 2015). Similarly, membership in microbiome communities should be influenced by direct selection of the genes in microbial genomes that confer a functional advantage to the microbe, such as those that respond to host diet. However, because genes are shared across microbial lineages, comparable community function may be achieved by different combinations of lineages, that is, communities may differ in taxonomic composition but be functionally similar. This has been observed in microbiomes across human body sites, which are broadly divergent in community composition, but appear to have highly conserved functional attributes (Human Microbiome Project 2012). Recognizing the relationship between genes and lineages would allow a role for metagenome function in the deterministic processes that may shape community composition, and would help explain high levels of variation common to datasets of host species (Shafquat et al. 2014).

Novel approaches that compare functional profiles and community composition from the same samples are needed to better resolve the relationship between structure and function and clarify assembly mechanisms. In this study, we progress this goal by characterizing how variance in community composition statistically relates to variance in predicted metagenome function. We develop our approach utilizing 16S amplicon sequencing due to the ability to link function and taxonomy to an extent that is currently challenged for complex communities using shotgun sequencing; however, the method can be applied to any dataset in which microbial taxonomy and function are jointly known. We provide an R package (FunkyTax; https://github.com/genotyper/FunkyTax) that includes functions to implement the novel components of this analysis workflow, as well as the data for this study. We focus on the gut microbiome of three insectivorous bat species. Bats exhibit the broadest range of dietary evolution among mammals over comparable timescales (Dumont et al. 2012), and distinct dietary strategies (e.g., insectivory, frugivory, sanguivory) are associated with consistent differences in microbiome composition (Phillips et al. 2012; Carrillo-Araujo et al. 2015). However, the majority of bat species are insectivorous and have evolved morphological and echolocation characteristics that enable fine-scale partitioning of the insect prey base (Denzinger and Schnitzler 2013). Among observed foraging strategies, insectivorous bats include fast-flying aerial hawking and slower-flying forest interior specialists. These diverse foraging strategies lead to differences in accessible insect prey species (Denzinger and Schnitzler 2013), providing a system in which host phenotypic differences select for modest and predictable dietary differences that may translate to selection differences on respective microbiomes. To characterize the interaction between microbiome structure and function, we used gut microbiome samples from three species of insectivorous bat from Kenya that differ in foraging strategy. We tested the hypothesis that metagenome variation can be explained by differences in dietary composition within a single trophic niche, and clarify the relationship between predicted metagenome function and microbiome community composition. We also discuss how metagenome functional differences among ecologically similar host species could be adaptive to respective ecologies.

Materials and methods

Collection and processing

Insectivorous bat species *Hipposideros beatus* (Hipposideridae) and *Kerivoula cuprosa* (Vespertilionidae) were captured at Kakamega forest, Kakamega County, Kenya. *Neoromicia tenuipinnis* (Vespertilionidae) was captured on the shores of Lake Victoria, Kisumu County, Kenya. The bat species *K. cuprosa* and *H. beatus* are forest interior foraging specialists, whereas *N. tenuipinnis* is a forest-edge aerial hawking species. Individuals were captured using harp traps and mist nets. Sex and

reproductive status were documented for each individual. Bats were held individually in new cloth holding bags until they defecated, then fecal material was collected and homogenized in RNALater (Life Technologies, Carlsbad, CA, USA). Total DNA was isolated from fecal samples using a MO BIO PowerMag Soil DNA Isolation Kit Optimized for Kingfisher (MO BIO Laboratories, Inc., Carlsbad, CA, USA) on an automated Kingfisher Flex platform (Thermo Scientific, Waltham, MA, USA).

16S amplification and sequencing

Samples were amplified for sequencing using a forward and reverse fusion primer encompassing variable regions 1 through 3 of the 16S gene. The forward primer was constructed with (5′-3′) the Illumina i5 adapter (AATGATACGGCGACCACCG AGATCTACAC), an 8–10 bp barcode, a primer pad, and 28F (GAGTTTGATCNTGGCTCAG; Wolcott et al. 2009). The reverse fusion primer was constructed with (5′-3′) the Illumina i7 adapter (CAAGCAGAAGACGGCATACGAGAT), an 8–10 bp barcode, a primer pad, and 519R (GAGTTTGATCN TGGCTCAG; Wolcott et al. 2009). Amplifications were performed in 25 μL reactions with Qiagen HotStar Taq master mix (Qiagen Inc., Valencia, CA, USA), 1uL of each 5uM primer, and 1uL of template. Reactions were performed on ABI Veriti thermocyclers (Applied Biosystems, Carlsbad, CA, USA) using the following thermal profile: 95°C for 5 min, then 35 cycles of 94°C for 30 sec, 54°C for 40 sec, 72°C for 1 min, followed by 1 cycle of 72°C for 10 min and 4°C hold. Amplification products were visualized with eGels (Life Technologies, Grand Island, New York). Products were pooled equimolar and each pool was size selected in two rounds using Agencourt AMPure XP (Beckman Coulter, Indianapolis, IN, USA) in a seven-tenths ratio of AMPure to product. Size selected pools were quantified using the Qubit 2.0 fluorometer (Life Technologies) and loaded on an Illumina MiSeq (Illumina, Inc. San Diego, CA, USA) 2 × 300 flow cell at 10 pM.

Cytochrome oxidase I amplification and sequencing

In order to maximize arthropod prey species detection, two separate arthropod Cytochrome oxidase I (COI) mini-barcode primer assays were employed. Library construction of these assays was performed in a series of two polymerase chain reaction (PCR) reactions. The first assay used MS_Art_1cF (AGATA TTGGAACWTTATATTTTATTTTTGG) and MS_ Art_2cR (WACTAATCAATTWCCAAATCCTCC) from Pons (2006). Thermal profile for first round PCR of this assay was 94°C for 3 min followed by 16 cycles of 94°C for 30 sec, 57°C for 30 sec (decreasing by 0.5C per cycle) and 72°C for 30 sec followed by 24 cycles of 94°C for 30 sec, 53°C for 30 sec and 72C° for 30 sec followed by a final extension at 72°C for 10 min. The second COI assay used MS_Lep1F (ATTCAACCAATCATAAAGATAT) and MS_LEP2R (CTTATATTATTTATTCGTGGGAAAGC) from Hebert et al. (2004). Thermal profile for first round PCR of this assay was 94°C for 1 min, 6 cycles of 94°C for 1 min, 45°C for 1 min 30 sec, 72°C for 1 min 15 sec, followed by 36 cycles of 94°C for 1 min, 51°C for 1 min 30 sec, and 72°C for 1 min 15 sec, followed by a final extension at 72°C for 5 min. For both assays, second round amplifications incorporated the Illumina i5 and i7 adapters, 8–10 bp barcodes, and used the same thermal profiles used in first-round reactions. All subsequent molecular methods were conducted same as described above.

Data processing

Sequence read pairs were stitched using PEAR (Zhang et al. 2014), and chimera-checking, operational taxonomic unit (OTU) clustering, community matrix development, and taxonomic assignment was conducted using standardized protocols described in Supplementary Information. All subsequent statistical analyses were conducted in R (Team 2015) using phyloseq (McMurdie and Holmes 2013), vegan (Oksanen et al. 2016), ape (Paradis et al. 2004), phytools (Revell 2012), and FunkyTax.

Community structure

Sequencing effort coverage was visually assessed using alpha diversity rarefaction curves of number of OTUs and phylogenetic diversity (PD; Faith 1992). Rarefaction was conducted with a step size of 250, between 250 and 15,000 classified reads, and 10 iterations at each step size. Differences in alpha diversity among bat species calculated from the full dataset were assessed using analysis of variance (ANOVA). Community compositional differences among species were summarized using both taxonomic (Bray–Curtis; Bray and Curtis 1957) and phylogenetic (UniFrac; Lozupone and Knight 2005) metrics. Inter-individual relationships based on resulting distance matrices were decomposed using non-metric multidimensional scaling non-metric multidimensional scaling (NMDS) and effects of species, sex, and the interaction of these variables was assessed using permutational analysis of variance using distance matrices (ADONIS) (Anderson 2001). Following a significant result for only host

species ($\alpha = 0.05$), post hoc ADONIS was repeated in a pairwise fashion between host species pairs. To further characterize microbiome community compositional differences among host species, the Euclidian distance of each individual from their respective group centroid resulting from NMDS was recorded and differences in group dispersion among bat species were assessed using ANOVA. Bacterial OTUs with significantly different abundances among host species were identified by assessing the effect of host species on OTU abundances. For this analysis, OTU count distributions were modeled using generalize linear models with a negative binomial error term as implemented in DESeq2 (Love et al. 2014). A false discovery rate of 5% was controlled using a Benjamini–Hochberg multiple testing correction (Benjamini and Hochberg 1995).

Functional prediction

A community matrix based on closed-reference OTU picking using the UCLUST algorithm against the Greengenes database (DeSantis et al. 2006) was constructed, and the count table subsequently corrected for 16S genomic copy number variation. A KEGG term-based (Kanehisa and Goto 2000) function matrix was developed from this matrix by comparing annotations of published bacterial genomes following the PICRUSt algorithm developed by Langille et al. (2013). Function was described at all KEGG annotation levels. Functional representation was assessed by linear regression comparing summary statistics including sum of function occurrences, number of unique functions, and number of reads. To gauge metagenomic predictive power, the nearest sequenced taxon index (NSTI; Langille et al. 2013) which summarizes weighted average divergence of observed bacterial lineages from those represented in the genomic database, was calculated for each sample. Compositional differences in species metagenomes were assessed using Bray–Curtis dissimilarities, ADONIS, and NMDS. Following a significant effect only for host species ($\alpha = 0.05$), post hoc ADONIS was repeated in a pairwise fashion between host species pairs.

Functional community categories

To clarify the relationship between structure and predicted function we characterized functional community categories by asking whether individual functions differed in frequency among host species, and then compared the community components contributing individual predicted functions. We conducted univariate tests for effect of host species on the abundance of each predicted function using generalized linear modeling as implemented in DESeq2 (Love et al. 2014). Next, for each function, matrices were constructed consisting of hosts (individual bats) by bacterial OTUs inferred to contribute a given function. To alleviate compositionality bias Hellinger data transformations (Legendre and Gallagher 2001) were performed on each matrix then host x host Bray-Curtis dissimilarities were calculated. Effect of host species on the composition of bacterial OTUs contributing to a predicted function was assessed with ADONIS. For all tests a false discovery rate of 5% was controlled using a Benjamini–Hochberg multiple testing correction (Benjamini and Hochberg 1995). The above data parsing and statistical tests can be implemented using R function TaFuR (available in package FunkyTax). Comparison of univariate and multivariate test results provided the basis of functional classification, illustrated in Fig. 1, and these comparisons can be made using R function *CatFun*. Briefly, this approach first assessed whether functions significantly differed in frequency across host species. It then determined whether non-significant functions were due to similarity in bacterial community composition among hosts (i.e., conserved community components), or contributed by homologous gene sets shared across different bacterial lineages (i.e., equivalent community components). Divergence in function correlated with divergence in taxonomic composition as expected in some cases (divergent community components), but divergent functions also resulted from abundance differences (i.e., enhanced community components).

To understand how the distribution of predicted functions across lineages may influence abundances and classifications to functional community categories, we quantified each predicted function's phylogenetic and taxonomic distribution. An overall bacterial phylogeny was estimated from full length Greengenes 16S sequences corresponding to observed OTUs, which were aligned using SSU-ALIGN (Nawrocki and Eddy 2010) and phylogeny estimated using FastTree2. Next, separate phylogenetic trees were created for bacterial OTUs contributing to each predicted function by pruning away from the overall phylogeny any bacterial OTUs not inferred to contribute to the frequency of each function. The sums of branch lengths were computed from resulting phylogenies. From this, the sum branch length contributing to each function was used as a measure of each function's phylogenetic distribution. A taxonomic component was provided by summarizing each function's contributing number of phyla. The relationship between function rank abundance and contributing phylogenetic branch length or phyla were summarized by

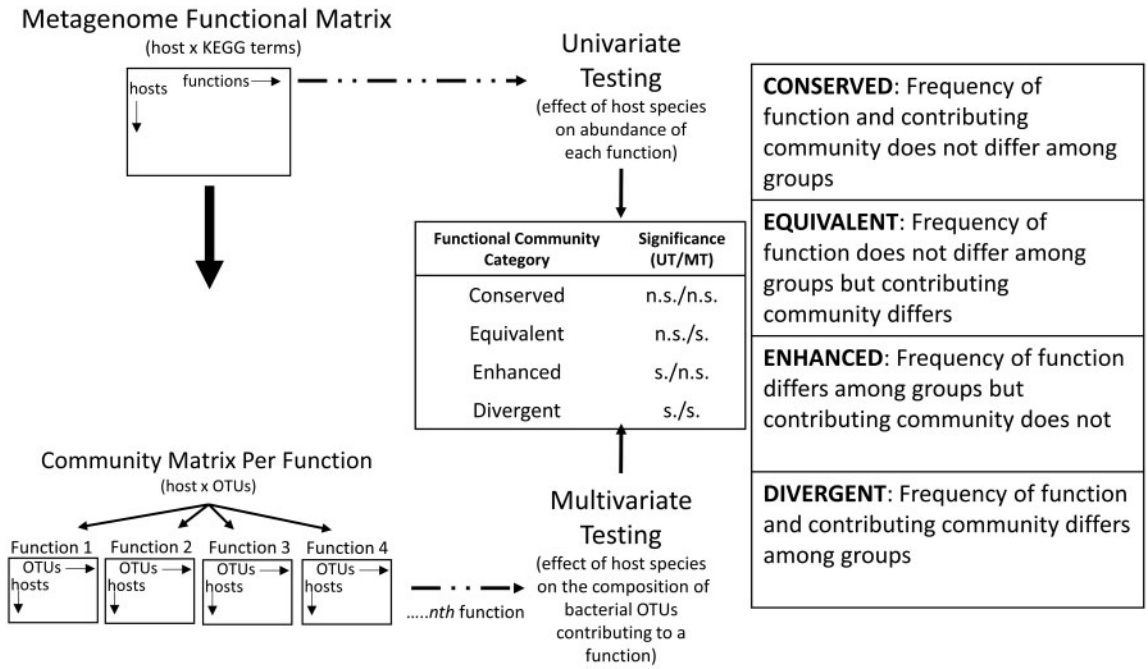

Fig. 1 Diagram illustrating the statistical testing and inferences used to classify microbiomes into functional community categories. An effect of a factor of interest (i.e., host species) on the abundance of each function is assessed through univariate testing (generalized linear modeling; Love et al. 2014). Because each metagenome function is contributed by a subset of the community, separate community matrices per function summarize how OTU composition contributing a function differs among individual hosts. Multivariate testing (ADONIS; Anderson 2001) is applied to each community matrix to test for an effect of a factor or interest (in this case host species) on composition (R function TaFuR). The results of both tests are used to classify functions (R function CatFun). Benjamini–Hochberg multiple testing correction (Benjamini and Hochberg 1995) is used to control the false discovery rate.

LOESS plots over a 10-function rank sliding window and statistically assessed using linear regression. Differences in the distribution of branch lengths for functions contributed by functional community categories (i.e., conserved, equivalent, divergent, enhanced communities; Fig. 1) were tested using ANOVA.

Molecular dietary analysis

A dietary data matrix was developed from COI arthropod barcode sequencing effort (see Supplementary Information). The dietary matrix was summarized from OTU through ordinal levels, and the ability of host species to explain differences in diet among individuals was assessed using ANOVA. Compositional difference in prey communities consumed across host species was assessed using Bray–Curtis dissimilarities, ADONIS, and NMDS. Following a significant effect only for host species on overall dietary compositional variation ($\alpha = 0.05$), post hoc ADONIS was repeated in a pairwise fashion between host species pairs.

Comparisons of microbiome community structure, predicted function, and host diet

To assess effects of host dietary composition on microbiome community structure and function, procrustes rotations were conducted. This approach was employed because the procrustean superimposition approach (Gower 1971), which compares ordination solutions rather than single distance measures, has been shown to be more powerful than Mantel testing over a range of scenarios (Peres-Neto and Jackson 2001). Rotations comparing microbiome community composition and diet, or metagenome function and diet, were based on results of distance-based redundancy analysis using Bray–Curtis dissimilarities, and separate comparisons were made with community composition and diet summarized at different taxonomic levels. Significance of procrustes rotations were assessed by comparison of observed residuals to null distributions obtained by randomizing samples in the rotated matrix through 1000 permutations (Jackson 1955) using function *protest* in the R package, vegan (Oksanen et al. 2016).

Results and Discussion

Effect of host species on microbiome diversity and community composition

Rarefaction curves of number of bacterial OTUs and Faith's PD as a function of sequencing effort

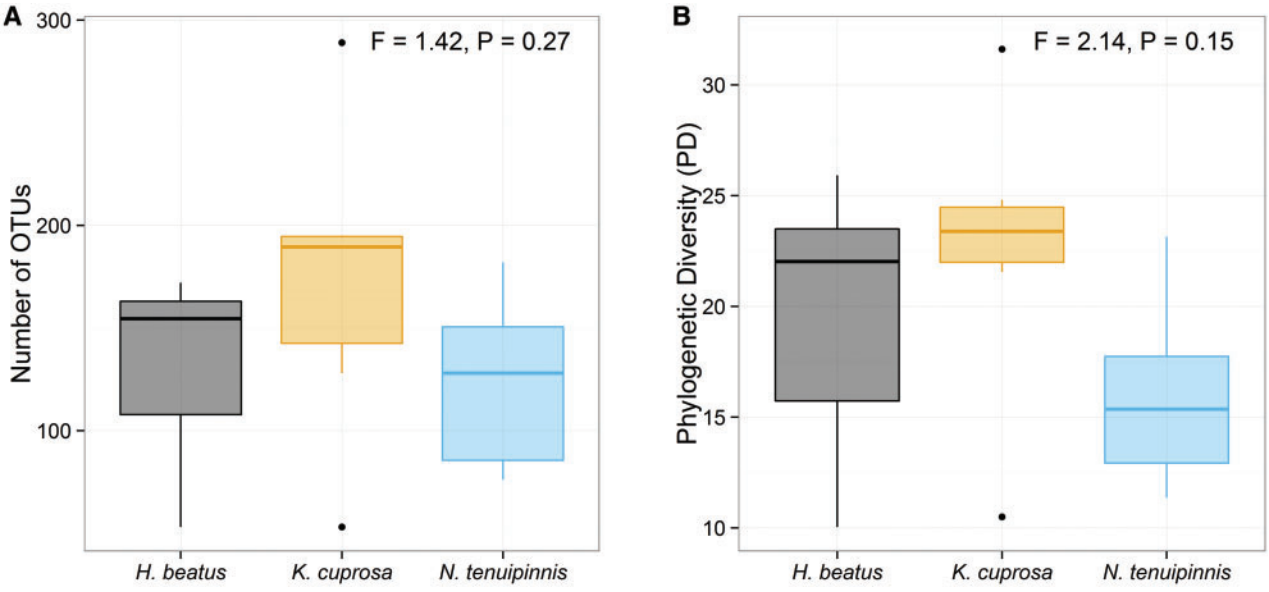

Fig. 2 Alpha diversity of microbiomes of three host species estimated as: (**A**) OTU richness; and (**B**) Faith's PD. F statistics and *P*-values are for ANOVA test of effect of bat species on alpha diversity estimate. *H* = *Hipposideros*, *K* = *Kerivoula*, *N* = *Neoromica*.

approached asymptote across samples, suggesting the estimated community matrix was not biased by poor sampling effort (Supplementary Fig. 1, Supplementary Table 1). Additionally, the distribution of observed OTUs was not significantly different from that estimated by Chao1 ($F = 0.64$, $P = 0.83$), an estimator that uses the frequency of singletons and doubletons to estimate missed diversity. No significant differences in unique OTUs or PD were observed among host species (Fig. 2A and B). Also, no differences in alpha diversity were found when the identified outlier (TK182019, χ^2 outlier test *P*-values 0.01 and 0.043 for OTUs and PD, respectively) was removed and comparisons were repeated (OTUs, $F = 0.48$, $P = 0.63$; PD, $F = 1.35$, $P = 0.29$). Variation in community composition among individuals and host species was characterized with Bray–Curtis and UniFrac measures of taxonomic and phylogenetic multivariate distance. Host species explained a significant proportion of the variation in the community matrix ($R^2 = 0.19$–0.24, Fig. 3A and B, Supplementary Table 2), and significance was generally observed when comparisons were made in a pairwise fashion between host species (Supplementary Table 2). In addition to compositional membership effects, significant group differences can also be generated by differences in levels of within-group variability. However, levels of inter-individual variation in community structure were similar across bat species (Fig. 3C and D), so we concluded that the primary difference in host species microbiome composition is related to which bacterial lineages occur across host species.

Because host species significantly affected community composition, we next evaluated bacterial OTU frequency differences in greater detail. Comparison of the number of shared and unique bacterial OTUs across host species revealed that *H. beatus* and *K. cuprosa*, the two ecologically similar forest interior foragers shared 30% of study-wide observed OTUs, while both shared 23% of all OTUs with the ecologically divergent *N. tenuipinnis* (i.e., 53 fewer common OTUs; Supplementary Table 3). When OTU abundances were considered in a univariate screen for group differences, the greatest number of significantly different OTUs were observed between *H. beatus* and *N. tenuipinnis*, the two host lineages that may be considered most different when phylogenetic distance and niche divergence are jointly considered. Significantly different OTUs were distributed across 11 bacterial phyla, with Firmicutes and Proteobacteria, the most commonly observed classifications (Supplementary Fig. 2).

Effect of host species on predicted functional capacity of the microbiome

We developed a metagenome functional matrix of hosts by KEGG terms following Langille et al. (2013) that comprised 5749 predicted functions. Linear comparisons of the number of reads, OTUs, and KEGG terms suggested that sequencing effort was sufficient to characterize metagenome functional

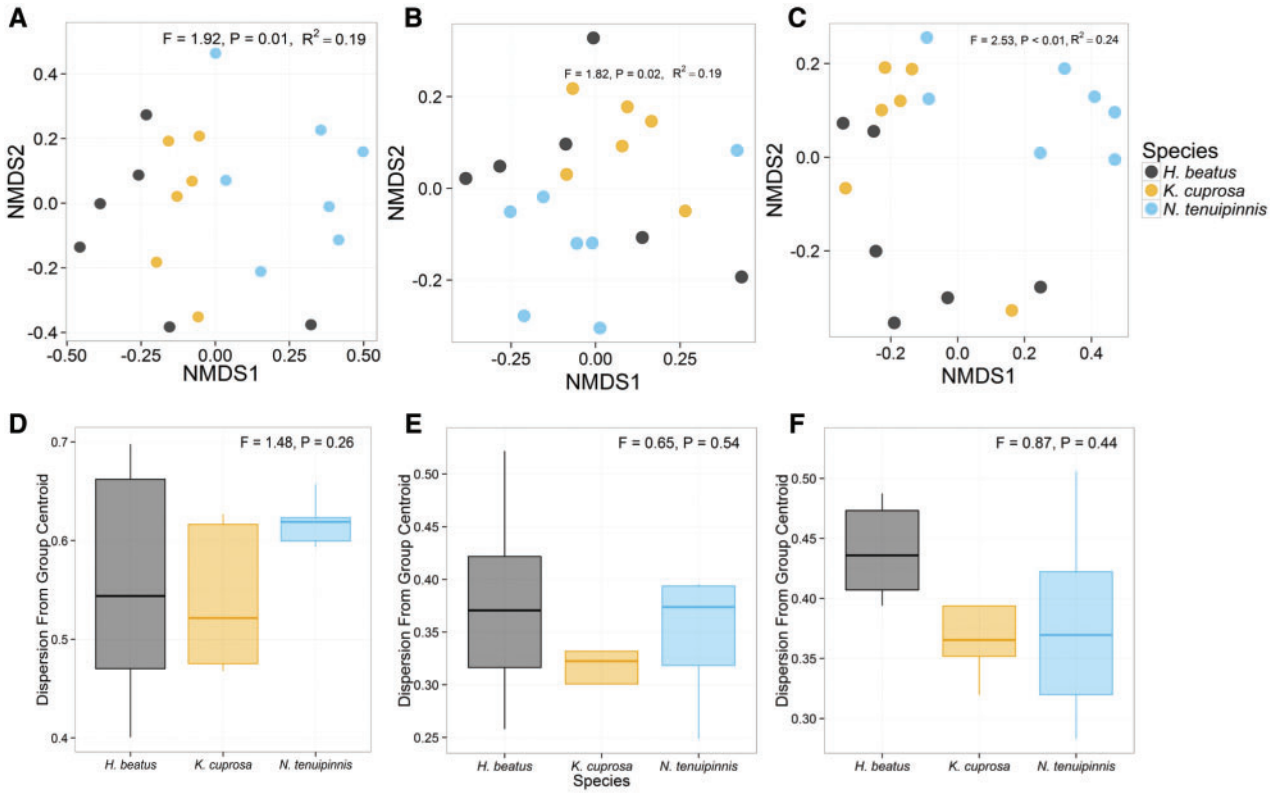

Fig. 3 Beta diversity of microbiomes of three host species estimated as: (**A**) NMDS based on Bray–Curtis dissimilarity index; (**B**) weighted UniFrac; or (**C**) unweighted UniFrac; and group dispersion based on (**D**) Bray–Curtis dissimilarity index; (**E**) weighted UniFrac; and (**F**) unweighted UniFrac. F statistics, R^2, and P- values for **A)** and **B**) are for ADONIS test of effect of bat species on beta diversity estimates, and F statistics and *P*-values for **C**) and **D**) are for ANOVA test for variation in group dispersion explained by bat species.

capacity (Supplementary Fig. 4). Values for the nearest sequenced taxon index (NSTI; Supplementary Fig. 4), a metric developed by Langille et al. (2013) to assess divergence of observed bacterial lineages from those represented in the database, were generally in the ranges observed for human microbiome samples. This result suggested good predictive power of the functional matrix, although NSTI values in this range are associated with variation in metagenome prediction accuracy (Fig. 3 of Langille et al. (2013)). There was an effect of host species on variation in the functional matrix that explained 21–23% of total variation, depending on whether or not function frequencies were considered (ADONIS, $F = 2.15$, $P < 0.01$ and $F = 2.44$, $P = < 0.01$, respectively). When functional variation was considered in a post hoc pairwise fashion among host species, mixed patterns of significance were observed (Supplementary Table 2).

Relationship between microbiome composition and predicted function

Hosts exhibited substantial variation in the relative abundance of dominant bacterial OTUs, but the proportions of dominant microbiome predicted functions were relatively consistent (Fig. 4), as in comparison of microbiome composition among human body sites (HMPC 2012). We hypothesized that the contrast between community structure and function could be partly explained by the distribution of genes across bacterial lineages. To provide insight into this relationship we first rank-ordered all 5749 predicted functions by frequency and then coded their occurrence by host species. Functions of higher rank-order (approximately the highest 50% of ranks) were generally similar in frequency of occurrence across species (Fig. 5A). In addition, we observed a strong linear relationship between functional rank-order and number of contributing bacterial phyla ($F = 7326$, $P < 0.01$), as well as functional rank-order and contributing bacterial phylogenetic branch lengths ($F = 8367$, $P < 0.01$; Supplementary Fig. 5A and B). These data indicated that the rank abundance of a given function was largely determined by its phylogenetic distribution. We further investigated this relationship by calculating the variance of each function's rank-order position across host species (Supplementary Fig. 5C). Positional variation was

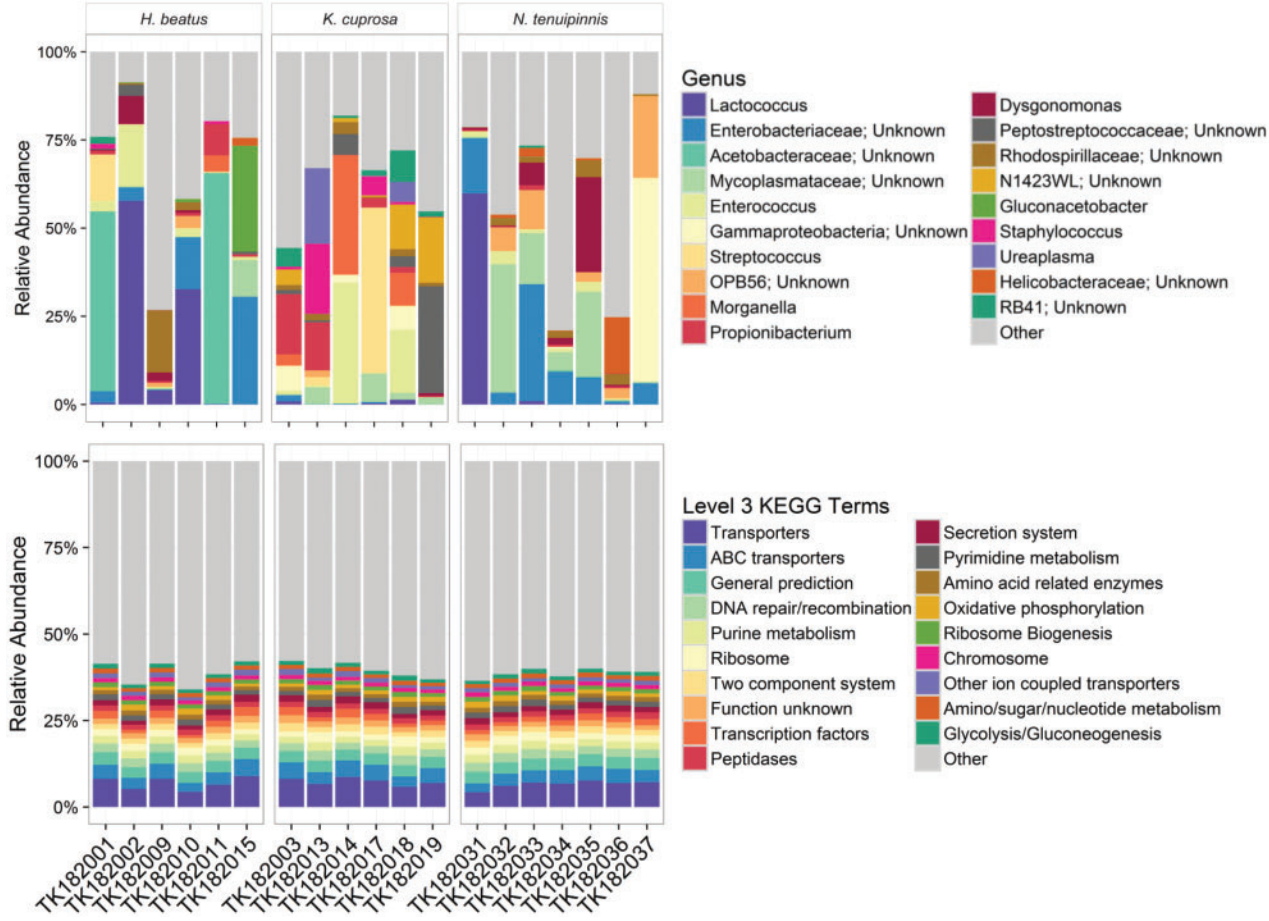

Fig. 4 Bar plots by individual host and grouped by host species of (**A**) relative abundance of top 20 most common bacterial genera (or reported at the lowest-level for which there was a confident assignment) and, (**B**) relative abundance of top 20 most commonly predicted metagenome functions (summarized as level 3 KEGG terms), in individual bat hosts.

smallest for the most and least frequent functions, with functions of intermediate overall rank-order displaying the most variation in rank-order among host species. Likely, the reduced rank-order variance of the most common functions is a result of their ubiquitous occurrence across all or most bacterial phyla and phylogenetic branch lengths (Supplementary Fig. 5A and B). Conversely, the reduced rank-order variance of the least common functions likely reflects their recent derivation (Fig. 5A).

Functional community categories

Because functional frequency trends across metagenomes are influenced by the pattern of shared and derived genes across microbial phylogeny, we developed an analytical framework to characterize how community structure contributes to conservation and divergence of predicted functions across hosts (Figs. 1 and 5B). We found that a large proportion (about three-quarters) of all functions have frequency distributions that do not differ significantly across host species. However, for the community components predicted to contribute these functions, we identified two classifications based on taxonomic compositions across host species. The first classification, "equivalent" community components, are those that contribute similar functional frequencies across hosts, but differ significantly in community composition. Functions arising from "equivalent" community components were disproportionately the most phylogenetically widespread (Fig. 5C), and exhibited conserved frequencies across hosts owing to the broad array of microbial lineages contributing these functions. The second classification, from "conserved" community components, were similar in both composition and function frequencies across hosts. Generally, "conserved" community components contributed functions occurring at the lower half of the rank-order distribution and were phylogenetically more derived than most 'equivalent' functions. We suggest that functions shared with similar frequencies across host species may represent the proposed core metagenome (Tettelin et al. 2005; Shafquat et al. 2014) among this set of host species.

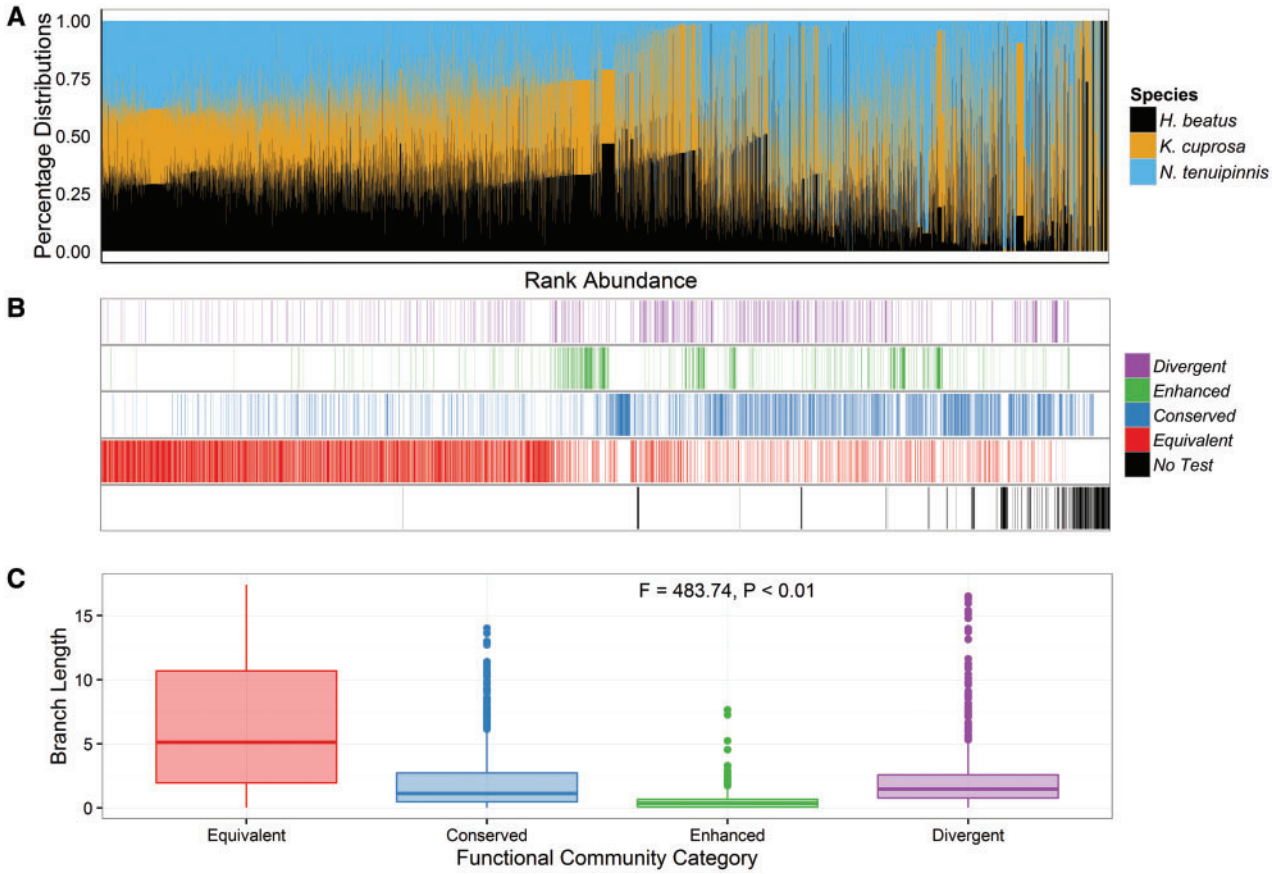

Fig. 5 Microbiome predicted function frequency distribution and corresponding functional community categories for each function: (**A**) base-level KEGG terms ordered from left to right by rank abundance and colored according to the percentage of each predicted function's occurrence in each bat species; (**B**) results of functional assignment to community categories; (**C**) distributions of phylogenetic branch length contributing to each function and summarized by community category. F statistic and *P*-value are results of ANOVA test for differences in branch length distribution among categories. See Supplementary Fig. 1 and text for details.

Predicted functions that significantly differed in frequency across host species may be influenced by host lineage-specific selection on community structure for specific functional characteristics. We were able to identify this group of functions and further characterize the community components contributing these functions. The first group, "enhanced" community components, did not significantly differ in taxonomic composition across hosts. Frequency differences in these functions were signified by equitable increases in the abundance of lineages contributing the functions. In comparison, "divergent" community components significantly differed in the community composition of the lineages contributing a given function. Notably, both enhanced and divergent functions were disproportionately distributed in the lower half of the rank-order distribution (Fig. 5B and C), and were contributed by fewer bacterial OTUs, phyla, and phylogenetic branch length (Supplementary Fig. 5). This suggests that differences in selection on metagenomes by host lineages with similar dietary ecologies may occur by selection for relatively derived metagenome functions. It should be noted that while "divergent" and "enhanced" functions disproportionately involved energetically-relevant traits (see below), relatively derived traits may also be more likely to experience lateral gene-transfer (Vos et al. 2015). Future assembly-based studies could help determine the relationship between transfer and functional divergence among environments.

Potential adaptive significance of enhanced and divergent functions

We categorized predicted functions that differed among hosts into their respective hierarchical KEGG categories. We found that "enhanced" and "divergent" functions contributing to significantly different level 3 KEGG terms among host species were disproportionately children of "Metabolic Processes" as opposed to all other level 1 KEGG categories ($\chi^2 = 4.31$, $P = 0.04$, Supplementary Fig. 6), suggesting that host lineage-specific metagenome

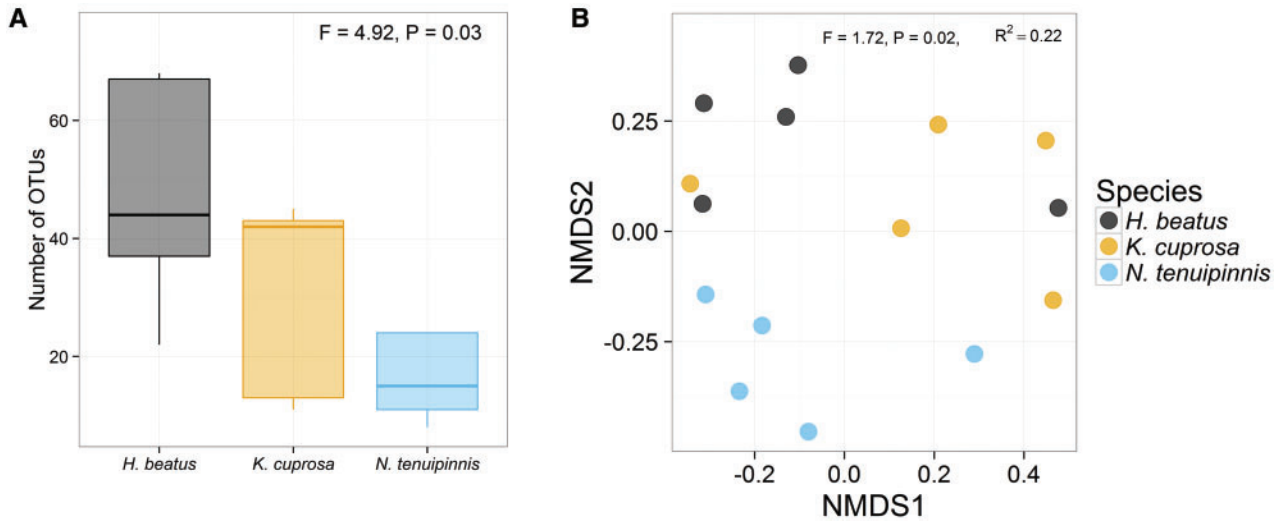

Fig. 6 Summary of dietary dataset for each host species with: (A) alpha diversity as number of unique OTUs (richness) represented in the diets of bat species. F statistics and *P*-values are for ANOVA test of effect of bat species on alpha diversity estimate; (**B**) Beta diversity as NMDS based on unweighted UniFrac. F statistics, R^2, and *P*-values are for ADONIS test of effect of bat species on beta diversity estimates.

divergence has occurred for genes relevant to energetic demands. Among metabolic functions with significant group differences there was considerable diversity in cellular processes. Some functions were not apparently relevant to differences among host species (e.g., flagellar assembly, bladder cancer, sporulation). The inclusion of such functions might be related to the genomic linkage of their genes to genomic regions that are under selection to increase metagenomic abundance, an incomplete annotation of their functions, or effects of the one-to-many relationship of genes to higher-level pathways. However, it is notable that significant functions included metabolism or biosynthesis of proteins, lipids, and carbohydrates. Given that comparisons were made among insectivorous host species, frequency differences for metagenome functions pertaining to macromolecule metabolism may reflect host lineage-specific metabolic fine-tuning. Metabolism or biosynthesis of several amino acids were different among hosts, including biosynthesis of essential amino acid lysine, and metabolism of essential nutrient ascorbate. The presence of arachidonic acid metabolism and fatty acid biosynthesis on this list may indicate an important role for metagenomes in sequestering cell membrane molecules, some of which are hypothesized to stabilize organs during body temperature reductions such as torpor (Ruf and Arnold 2008). Also, a significant increase in lipid metabolic functions were inferred for *K. cuprosa* relative to *N. tenuipinnis*. Utilization of intrinsic lipid metabolic pathways of insectivorous bats has been hypothesized as a mechanism for meeting energetic demands of volant flight (Voigt et al. 2010; McGuire et al. 2013; Phillips et al. 2014). Current data support a role for metagenomes in bat-specific lipid requirements.

Host diet, microbiome composition, and function

We hypothesized that host dietary ecology is a primary selective pressure on metagenome function and characterized host diet using mitochondrial COI arthropod barcode amplicon sequencing (Hebert et al. 2004; Pons 2006). A weak relationship was observed between classified COI reads and OTUs ($F = 3.84$, $P = 0.07$), indicating an effect of sampling effort on inferred diets. However, there was no significant difference in sampling effort among host species ($F = 0.91$, $P = 0.43$), which indicated that differences in dietary composition among host species was not influenced by sampling effort. There was variation in the number of OTUs consumed across host species; forest interior specialists on average consumed more OTUs, with *H. beatus* consuming the most, and *N. tenuipinnis* (aerial hawking species) consuming the lowest prey diversity (Fig. 6A). In addition, a significant proportion of the dietary matrix was explained by host species ($R^2 = 0.22$, Fig. 6B). As expected given the contrasts in host dietary niche, post hoc pairwise testing revealed significant dietary differences between *N. tenuipinnis* and both forest interior host species (*H. beatus* and *K. cuprosa)*, whereas the diets of these latter two did not significantly differ (Supplementary Table 2).

Because community composition, predicted function, and dietary datasets all generally indicated the

largest divergences occurred between the aerial hawking species (*N. tenuipinnis*) and either of the two forest interior specialists (*H. beatus*, *K. cuprosa*), we assessed the congruence of community and functional datasets to the dietary dataset using procrustean superimposition (Jackson 1955; Peres-Neto and Jackson 2001). Because effects of dietary composition on community and functional variation may vary with taxonomic level, we summarized community and dietary datasets at multiple taxonomic levels, and performed procrustean rotations for each rank. Supporting a direct effect of dietary differences among hosts on metagenome function, the overall strongest correlation was observed between the dietary matrix summarized at prey species-level and the predicted functional matrix (Procrustean Correlation = 0.52, $P = 0.03$, Fig. 7A, Supplementary Table 4). There was a trend in strength of correlation across dietary taxonomy, with a decay occurring as diet was summarized at higher taxonomic ranks, suggesting that dietary effects distinguishing metagenome functional profiles across host species are determined by nutritional differences arising at the level of prey species or genus. In contrast, no significant correlations were observed between the community and dietary matrices summarized at any taxonomic level (Fig. 7B, Supplementary Table 5). These results suggest that selection by host diet acts on metagenome function, and only secondarily on bacterial lineages. That is, because genes are shared across bacterial lineages, multiple bacterial lineages can provide the same functions and are effectively interchangeable. Lineage interchangeability dilutes the directly measurable response of microbiome community composition to selection for metagenome function.

Direct selection for metagenome function in combination with the distribution of functions across microbiome communities contributes to complex patterns describing relationships between hosts and microbiomes. In this study we characterized how microbiome communities may dynamically change, and how compositional variation relates to predicted function frequency across hosts. We applied our analytical approach to understand metagenome structure–function relationships among similar host dietary ecologies, but our approach can be applied to a range of environments with general predictions described in Supplementary Fig. 7. In the present study we made metagenome inferences from 16S amplicon data, an approach unable to incorporate information on gene gain/loss and lateral gene transfer specific to observed lineages; however, this did not preclude detection of statistical signal for diet–function relationships that was more evident than diet–community structure

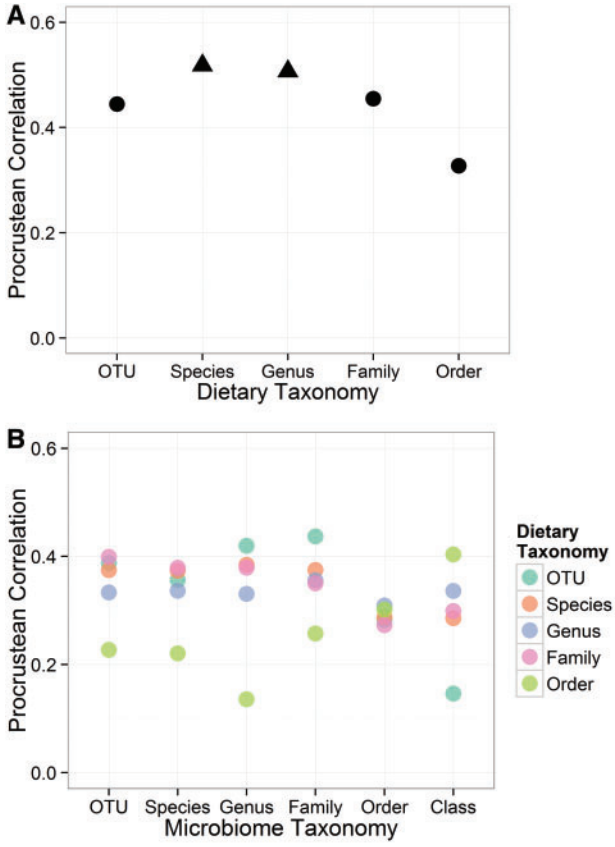

Fig. 7 Correspondence between bat diet and microbiome predicted function and taxonomy. (**A**) Procrustean correlations of the functional matrix to the dietary matrix summarized at different taxonomic levels. (**B**) Procrustean correlations of the community and dietary matrices, both summarized at different taxonomic levels. Triangles represent correlations that were identified as significant through permutation.

relationships. As the challenges associated with using shotgun sequencing for comprehensive taxonomic and functional characterization diminish, the approach presented here will aid in the integration of genome evolution and community ecology.

Supplementary data

Supplementary Data available at *ICB* online.

Data accessibility

The function matrices, community matrices, and metadata associated with this study are available as example data with R package FunkyTax, available at https://github.com/genotyper/FunkyTax. Sequence data analyzed in this study is deposited in SRA under accession number [PRJNA383585].

Acknowledgments

We would like to thank the Natural Science Research Laboratory at Texas Tech University for field

supplies, sample archival, and loan approval, Liam Mcguire and Robert Bradley for helpful comments on an earlier draft, Dylan Schwilk for programing expertise, RTL Genomics for financial, laboratory, and bioinformatic support, and the Kenya Wildlife Service for granting access and research permits.

Funding

T.K. was supported by a travel grant from Texas Tech University's Office of International Affairs, College of Arts and Sciences, and the Department of Biological Sciences. P.W. was supported by an Overseas grant from the British Ecological Society and the International Foundation of Science. This work was supported by the National Science Foundation [grant number IOS-1638630].

References

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austal Ecol 26:32–46.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 57:289–300.

Bolnick DI, Snowberg LK, Hirsch PE, Lauber CL, Knight R, Caporaso JG, Svanback R. 2014. Individuals' diet diversity influences gut microbial diversity in two freshwater fish (threespine stickleback and Eurasian perch). Ecol Lett 17:979–87.

Bray JR, Curtis JT. 1957. An ordination of upland forest communities of southern Wisconsin. Ecol Monogr 27:325–49.

Carrillo-Araujo M, Tas N, Alcantara-Hernandez RJ, Gaona O, Schondube JE, Medellin RA, Jansson JK, Falcon LI. 2015. Phyllostomid bat microbiome composition is associated to host phylogeny and feeding strategies. Front Microbiol 6:447.

Denzinger A, Schnitzler HU. 2013. Bat guilds, a concept to classify the highly diverse foraging and echolocation behaviors of microchiropteran bats. Front Physiol 4:164.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–72.

Dumont ER, Davalos LM, Goldberg A, Santana SE, Rex K, Voigt CC. 2012. Morphological innovation, diversification and invasion of a new adaptive zone. Proc Biol Sci 279:1797–805.

Faith DP. 1992. Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1–10.

Gower J. 1971. Statistical methods of comparing different multivariate analyses of the same data. Mathematics in the archaeological and historical sciences. Edinburgh University Press. p. 138–49.

Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci U S A 101:14812–17.

Hollister EB, Gao C, Versalovic J. 2014. Compositional and functional features of the gastrointestinal microbiome and their effects on human health. Gastroenterology 146:1449–58.

Hubbell SP. 2006. Neutral theory and the evolution of ecological equivalence. Ecology 87:1387–98.

Human Microbiome Project C. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–14.

Jackson D. 1955. PROTEST: A PROCrustean Randomization TEST of community environment concordance. Ecoscience. 2:297–303.

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

Kumar PS, Mason MR. 2015. Mouthguards: does the indigenous microbiome play a role in maintaining oral health?. Front Cell Infect Microbiol 5:35.

Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, et al. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31:814–21.

Lathrop SK, Bloom SM, Rao SM, Nutsch K, Lio CW, Santacruz N, Peterson DA, Stappenbeck TS, Hsieh CS. 2011. Peripheral education of the immune system by colonic commensal microbiota. Nature 478:250–4.

Legendre P, Gallagher E. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129:271–80.

Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, et al. 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett 7:601–13.

Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, et al. 2008. Evolution of mammals and their gut microbes. Science 320:1647–51.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–35.

McGuire LP, Fenton MB, Guglielmo CG. 2013. Phenotypic flexibility in migrating bats: seasonal variation in body composition, organ sizes and fatty acid profiles. J Exp Biol 216:800–8.

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

Muegge BD, Kuczynski J, Knights D, Clemente JC, Gonzalez A, Fontana L, Henrissat B, Knight R, Gordon JI. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science 332:970–4.

Nawrocki EP, Eddy SR. 2010. ssu-align: a tool for structural alignment of SSU rRNA sequences (http://selab. janelia. org/software. Html).

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2016. vegan: Community Ecology Package. R package version 2.3-4 (https://CRAN.R-project.org/package=vegan).

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–90.

Peres-Neto P, Jackson D. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129:169–78.

Phillips CD, Phelan G, Dowd SE, McDonough MM, Ferguson AW, Delton Hanson J, Siles L, Ordonez-Garza N, San Francisco M, Baker RJ. 2012. Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. Mol Ecol 21:2617–27.

Phillips CJ, Phillips CD, Goecks J, Lessa EP, Sotero-Caio CG, Tandler B, Gannon MR, Baker RJ. 2014. Dietary and flight energetic adaptations in a salivary gland transcriptome of an insectivorous bat. PLoS One 9:e83512.

Pons J. 2006. DNA-based identification of preys from non-destructive, total DNA extractions of predators using arthropod universal primers. Mol Ecol Notes 6:623–6.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–23.

Ricklefs RE, Travis J. 1980. A morphological approach to the study of avian community organization. Auk 97:321–38.

Round JL, Mazmanian SK. 2009. The gut microbiota shapes intestinal immune responses during health and disease. Nat Rev Immunol 9:313–23.

Ruf T, Arnold W. 2008. Effects of polyunsaturated fatty acids on hibernation and torpor: a review and hypothesis. Am J Physiol Regul Integr Comp Physiol 294:R1044–52.

Schmidt VT, Smith KF, Melvin DW, Amaral-Zettler LA. 2015. Community assembly of a euryhaline fish microbiome during salinity acclimation. Mol Ecol 24:2537–50.

Shafquat A, Joice R, Simmons SL, Huttenhower C. 2014. Functional and phylogenetic assembly of microbial communities in the human microbiome. Trends Microbiol 22:261–6.

Sommer F, Stahlman M, Ilkayeva O, Arnemo JM, Kindberg J, Josefsson J, Newgard CB, Frobert O, Backhed F. 2016. The gut microbiota modulates energy metabolism in the hibernating brown bear *Ursus arctos*. Cell Rep 14:1655–61.

Team RC. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (https://www.R-project.org/).

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, et al. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A 102:13950–5.

Turnbaugh PJ, Gordon JI. 2009. The core gut microbiome, energy balance and obesity. J Physiol 587:4153–8.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project. Nature 449:804–10.

Voigt CC, Sorgel K, Cechmann SKN. 2010. Refueling while flying: Foraging bats combust food rapidly and directly to power flight. Ecology 91:2908–17.

Vos M, Hesselman MC, te Beek TA, van Passel MW, Eyre-Walker A. 2015. Rates of lateral gene transfer in prokaryotes: high but why? Trends Microbiol 23:598–605.

Wolcott RD, Hanson JD, Rees EJ, Koenig LD, Phillips CD, Wolcott RA, Cox SB, White JS. 2016. Analysis of the chronic wound microbiota of 2,963 patients by 16S rDNA pyrosequencing. Wound Repair Regen 24:163–74.

Wolcott RD, Gontcharova V, Sun Y, Dowd SE. 2009. Evaluation of the bacterial diversity among and within individual venous leg ulcers using bacterial tag-encoded FLX and titanium amplicon pyrosequencing and metagenomic approaches. BMC Microbiol 9:226.

Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–20.