

# Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes

Patricio Jeraldo<sup>\* †</sup>, Maksim Sipos<sup>\* †</sup>, Nicholas Chia<sup>\* †</sup>, Jennifer M. Brulc<sup>\* ‡</sup>, A. Singh Dhillon<sup>§</sup>, Michael E. Konkel<sup>¶</sup>, Charles L. Larson<sup>¶</sup>, Karen E. Nelson<sup>||</sup>, Ani Qu<sup>\* ‡ \*\*</sup>, Lawrence B. Schook<sup>\* ‡</sup>, Fang Yang<sup>\* ‡</sup>, Bryan A. White<sup>\* ‡</sup>, and Nigel Goldenfeld<sup>\* †</sup>

<sup>\*</sup>Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, IL 61801, USA, <sup>†</sup>Loomis Laboratory of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801, USA, <sup>‡</sup>Department of Animal Sciences, 1207 West Gregory Drive, Urbana, IL 61801, USA, <sup>§</sup>Washington State University Avian Health and Food Safety Laboratory, Washington State University, Puyallup, WA 98371, USA, <sup>¶</sup>School of Molecular Biosciences, Washington State University, Pullman, WA 99164, USA, <sup>||</sup>The J. Craig Venter Institute, 9712 Medical Center Drive, Rockville, MD 20850, USA, and <sup>\*\*</sup>Environmental and Occupational Health Sciences Institute, 170 Frelinghuysen Road, Piscataway, NJ 08854, USA

Submitted to Proceedings of the National Academy of Sciences of the United States of America

**The theoretical description of the forces that shape ecological communities focus around two classes of models. In niche theory, deterministic interactions between species, individuals and the environment are considered the dominant factor, whereas in neutral theory, stochastic forces, such as demographic noise, speciation and immigration are dominant. Species abundance distributions predicted by the two classes of theory are difficult to distinguish empirically, making it problematic to deduce ecological dynamics from typical measures of diversity and community structure. Here we show that the fusion of species abundance data with genome-derived measures of evolutionary distance can provide a clear indication of ecological dynamics, capable of quantifying the relative roles played by niche and neutral forces. We apply this technique to six gastrointestinal microbiomes drawn from three different domesticated vertebrates, using high resolution surveys of microbial species abundance obtained from carefully curated deep 16S rRNA hypervariable tag sequencing data. Although the species abundance patterns are seemingly well fit by the neutral theory of metacommunity assembly, we show that this theory cannot account for the evolutionary patterns in the genomic data; moreover our analyses strongly suggest that these microbiomes have in fact been assembled through processes that involve a significant non-neutral (niche) contribution. Our results demonstrate that high-resolution genomics can remove the ambiguities of process inference inherent in classical ecological measures, and permits quantification of the forces shaping complex microbial communities.**

microbial ecology | niche theory | neutral theory | species abundance distribution | metagenomics

Abbreviations: OTU, operational taxonomic unit

**E**cological species distributions are determined by the interplay between environmental factors and evolutionary processes. In classical ecological theory, niches, characterized, for example, by nutrients and other environmental factors, determine species abundance distributions and populations primarily through deterministic partitioning of resources amongst species (1). Species populations are limited by niche carrying capacity, rather than interspecies competition, thus tending to promote coexistence (2). In niche theory, diversity is determined primarily by the number of available niches, raising the issue of how to account quantitatively for the apparent observed diversity (3–6) from well-documented instances of niche differences (7).

An alternative perspective is the class of neutral theories, in which species are functionally equivalent, and stochastic factors such as immigration, birth-death processes and speciation are the primary drivers of ecological diversity and community structure (8–13). This class of models has been reported to be capable of accurate predictions for the species abundance distributions in (e.g.) riverine

fish populations (14) or microbial populations (15), in addition to the early successes in forest ecosystems, a planktonic copepod community, and a bat community in Barro Colorado Island (BCI) (10). However, the methodology used in such comparisons is contentious when examined carefully (16, 17), with sampling issues, parameter estimation, and model definition being some of the key factors that require careful attention. The assumptions of neutral theory, in particular functional equivalence, are not transparently biological (18), and additionally have been criticized on a variety of empirical grounds (19, 20), including the predictions for species lifetimes, speciation rates and the incidence of rare species (21). Other technical assumptions, for example that the number of individuals competing for a resource is a constant (the “zero-sum” assumption), may be unrealistic, but can be extended or relaxed (13, 22, 23). Perhaps a more useful insight into the applicability of neutral theory comes from considering the interplay between niche stabilization mechanisms and fitness (24). A recent study of a sagebrush steppe community, where strong niche stabilization mechanisms were identified even in the presence of apparently small fitness differences (25), underscores the fact that weak functional inequivalence need not necessarily mean that niche dynamics are negligible. On the other hand, a study that attempted to infer pairwise interaction strengths among the most abundant species in the BCI site found that interspecies interactions were much weaker than intraspecies one, in apparent agreement with neutral assumptions (26).

Despite their fundamental differences, and the plethora of studies nominally supporting each side of the niche-neutral dichotomy, these theories predict species abundance distributions that are difficult to distinguish empirically (5, 27), with similar mathematical properties for asymptotically large diversity (28). The inverse problem of inferring ecological dynamics from measures of diversity does not appear to have a unique solution, either theoretically or empirically. Accordingly, a more nuanced perspective has arisen (2, 19, 29), in which elements of both types of theory may contribute to a proper description of the ecological dynamics, and a variety of mathematical frameworks for accomplishing this type of synthesis have recently

Reserved for Publication Footnotes

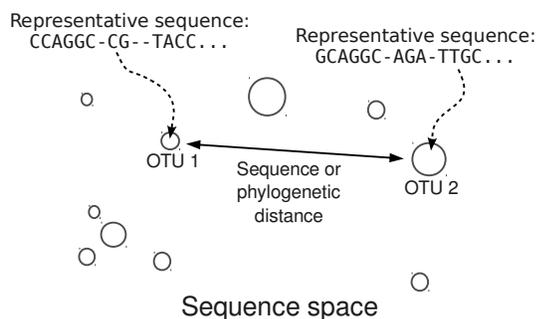
appeared (26,30–35). Nevertheless, it remains an open question as to how to properly characterize community dynamics, and how to usefully quantify the relative roles of niche and neutral processes in the evolutionary dynamics of ecosystems.

These questions are of particular relevance to microbial communities, which play functionally important roles in ecosystems, but are typically rich in diversity, suggesting the presence of sub-populations shaped primarily by stochastic forces. Such communities would not be expected to represent endmembers of the niche-neutral continuum, and quantification of their structuring process represents a complex problem that has recently attracted attention. Most studies find evidence for a mixture of neutral and niche processes in microbial community assembly (36–40). These seem to arise for different physical reasons. One indication is that the neutrally-assembling taxa are generalist microbes, that can exist in a wide variety of environments (38), whereas the niche portion of the microbiota are adapted to the media conditions (41). There are also indications that that microorganismal cooccurrence patterns are shaped by the same processes and interactions that shape macroorganismal cooccurrence patterns (42).

In this paper we propose a methodology for addressing the problem of quantifying the relative role of niche and neutral processes in structuring microbial communities, by fusing measures of abundance with phylogenetic information. The merging of classical ecological measures with phylogenetic analysis is growing in importance, but is still in its infancy (43–47). The method presented here is particularly applicable to uncultured microbial communities that are characterized by a high level of diversity, and are amenable to modern metagenomic tools, such as pyrosequencing.

In order to explain the basic idea of how we quantify an ecosystem on the niche-neutral continuum, it is necessary to recall how microbiomes can be probed by genomic methods. The first step in an ecological study of a microbiome, following sequencing, cleanup and alignment, is the assignment of sampled sequences into Operational Taxonomic Units (OTUs) through a clustering process (48). The OTUs are then used as a proxy for estimating microbial species abundance (49). The OTU data are two-fold. On the one hand, the OTUs have relative abundances that are estimations of the species' abundances in the environment. On the other hand, OTUs also have representative sequences associated with them. Typically a representative sequence of an OTU is the most abundant of the identical clones within the OTU, and also it is more than 97% similar to every other sequence within that OTU. This genomic data associated with the representative sequence allows us to think of OTUs as points in a sequence space as illustrated in Fig. 1. We can think of distances between points in this space as corresponding to the phylogenetic or sequence distances between the sequences in these OTUs.

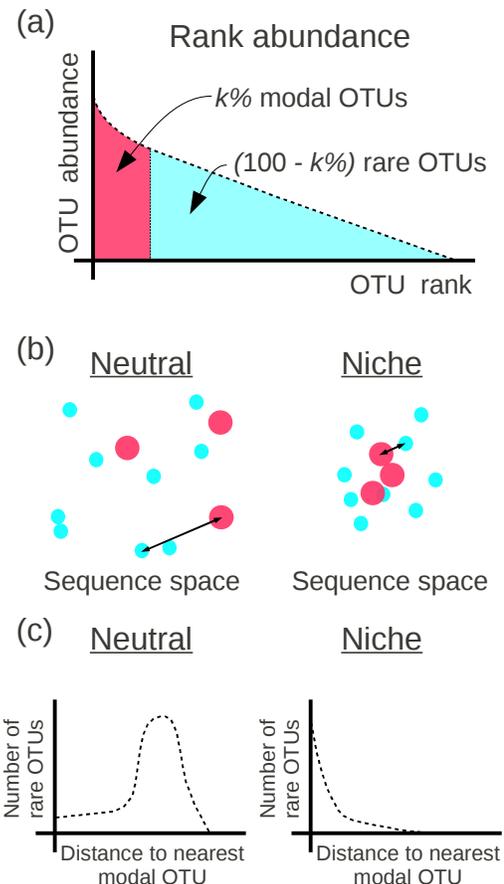
This cloud of points in high-dimensional sequence space can also be labeled by OTU abundance. In our work, this is determined by se-



**Fig. 1.** Sketch of the starting point for a metagenomic analysis of an environment. Circles indicate OTUs, and abundance (number of sequences within the OTU) is labeled by the size of the circle. A representative sequence is associated with each OTU. The OTUs are embedded in a sequence space such that the distance between the circles in the sequence space corresponds to e.g. sequence or phylogenetic distance between the representatives.

quence abundance (after every effort has been made to account for artifacts), but in principle OTU abundance labels could be obtained from any other source, such as Q-PCR. In this space, we can categorize the OTUs into two sorts: the most abundant OTUs (which we term “modal” OTUs, and define this precisely below) and the other, less abundant, OTUs (which we term “rare” OTUs, and define this precisely below). The correlations between the modal and rare OTUs will depend upon the evolutionary dynamics, and in fact exhibit sharp mathematical differences that can be used to discriminate different putative dynamics. To see the essential idea, we will now explain how this would work in two caricatures of ecosystem dynamics: a simplified neutral model and a simplified niche model. A significantly more elaborate analysis is carried out below, in the main body of this paper, but the key concepts are captured by these simplified models.

First, suppose that the evolutionary dynamics is itself neutral, so that the rare and modal OTUs are distributed at random in the high-dimensional sequence space. We are going to be interested in measuring the distances between sequences corresponding to different OTUs, and comparing their similarity. Let us assume that the sequences being analyzed are all of the same length, containing  $L$



**Fig. 2.** (a) Classification of the OTUs into two groups based on the rank abundance. The top  $k\%$  of OTUs are labeled modal, whereas the remainder of the OTUs are labeled rare. (b) Sketch of the neutral and niche evolution processes in sequence space. Light blue OTUs are rare, whereas red OTUs are modal. For the neutral process, the average distance of a rare OTU to its closest modal OTU is large (indicated by the arrow). For the niche process, this distance is much smaller since rare OTUs cluster about the modal OTUs which define the niches. (c) Sketch of the expected distributions of distance to the closest modal OTU. For the neutral process, this distribution is peaked around some non-zero distance, which is close to the average distance between the OTUs in the dataset. In the niche process, the distribution monotonically decays with distance since the rare OTUs are attracted to the niches.

nucleotide bases from the usual 4-letter alphabet (ACGT); here we are ignoring real life complications such as insertions, deletions and gaps. We label the sequences by  $S_{\alpha}^i$ , where  $\alpha = 1 \dots L$  labels position along the sequence and  $i$  labels the OTU;  $S_{\alpha}^i$  can take the values 1,2,3,4 corresponding to the alphabet of bases ACGT. We define the normalized Hamming distance  $H_{ij}$  between two sequences  $i$  and  $j$  as the fraction of bases in  $i$  that are different from the base in the corresponding position in  $j$ :

$$H_{ij} \equiv \frac{1}{L} \sum_{\alpha=1}^L (1 - \delta(S_{\alpha}^i - S_{\alpha}^j)) \quad [1]$$

where  $\delta$  denotes the Kronecker delta. The mean  $\langle H \rangle$  of  $H_{ij}$  averaged over a large sample of random sequences would be  $3/4$ , because there is a  $1/4$  chance that two bases at the same position are identical. Thus, the probability distribution of  $H$  would be expected to be a roughly bell-shaped curve, peaked around  $H = 3/4$ , with a width dependent on the number of sequences. In practice, there are complications due to insertions, deletions and gaps, but most importantly, conserved positions. Bases that are highly conserved cannot be appropriately modeled as being chosen randomly from the alphabet. This can be taken into account by simply restricting the above analysis to bases that are strongly non-conserved: let us call the number of highly conserved bases  $M < L$ , so that the expected value of  $H$  will now be reduced by the fraction of conserved bases:  $\langle H \rangle = 3(L - M)/4L$ . Thus, taking into account conservation, the bell-shaped curve will shift its peak to a smaller value of  $H$ . In the data presented below, we found that  $L \sim 200$  and  $M \sim 160$ , so that the distribution of  $H$  should be peaked at about  $0.15$ , in the case of a neutral system. Now consider a subset  $\{E_k\}$  of distances  $\{H_{ij}\}$ . For each “rare” OTU  $k$ , we rank all of the distances between OTU  $k$  and each “modal” OTU  $l$ . Then, we select the shortest such distance and label it  $E_k$ . In this way, the set  $\{E_k\}$  is the set of distances of “rare” OTUs to their nearest niche neighbor. For the above case where the evolutionary dynamics is neutral-like, the distribution of  $E$  is also a bell-shaped curve like the distribution of  $H$ . However, its mean is slightly shifted to the smaller values, and its standard deviation is smaller (because  $\{E\}$  is the subset of shortest distances from the set of  $\{H\}$ ). In other words,  $\langle E \rangle < \langle H \rangle$ .

Second, let us consider a caricature of a system that is dominated by niche dynamics. In the extreme (and unrealistic) case where there is only one niche, occupied by one particular modal OTU, the probability distribution of  $E$  will be a delta distribution peaked at  $E = 0$ . In a more realistic model, where there is a cloud of rare OTUs surrounding the modal OTU, having evolved from it by a few point mutations, one would expect the probability distribution of  $E$  to be peaked at  $E = 0$ , and then to monotonically decrease for  $E > 0$ . In the case of a system with several niches, the probability distribution for  $E$  will be somewhat more complicated, because one needs to calculate the normalized Hamming distance from each rare OTU to the nearest modal OTU, and this requires making a Voronoi polyhedron construction in sequence space. Nevertheless, for small values of  $E$ , the probability distribution will be dominated by the single niche argument given above, and the functional form will be unchanged: peaked at the origin and monotonically decreasing for  $E > 0$ . These two caricatures for simplified models of ecosystem structure are sketched in Fig. 2, and show that there are clear and distinct signatures arising from the nature of the processes that have structured the community.

In the remainder of this paper, we numerically evaluate the metric for model systems in order to quantitatively and concretely confirm the above heuristic description. We then describe how we have implemented these ideas in a proof-of-principle study of vertebrate gastrointestinal microbiomes. These experimental systems were chosen, not only because of the growing recognition of the importance of mi-

crobiomes as a determinant of host health (50), but also because these are systems that have high diversity, and are likely to be shaped both by stochastic and niche processes. Indeed, as we will see, they can be well-described naively by neutral theory, although in fact niche processes play a fundamental role in structuring these communities.

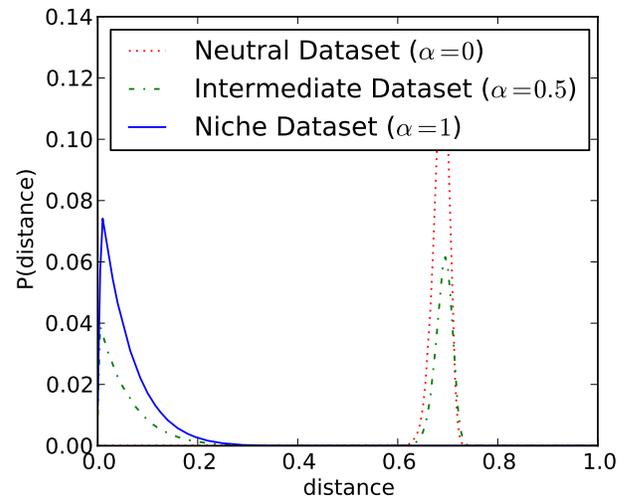
## Model calculations

In this section we evaluate our metric on model systems parametrized by a single parameter,  $\alpha$ , the proportion of the system undergoing a niche dynamic. We perform 5000 Monte Carlo simulations of the following process. We simulate  $N$  OTUs (here  $N = 1000$ ) each with representative sequences of length  $L = 200$ . A subset  $\alpha N$  ( $0 \leq \alpha \leq 1$ ) of the OTUs undergo a niche dynamic in the following way. A single random OTU is chosen to be the center of the niche. The remainder of the  $\alpha N - 1$  OTUs (niche OTUs) are generated by performing random mutations of the genome of the OTU representing the niche center. The placement and number of the mutations were chosen randomly in the following way. Placements of mutations were sampled uniformly (without replacement) across the entire genome. The number of mutations for each of the niche OTUs was sampled from an exponential distribution thereby modeling the evolution of OTUs under multiplicative fitness pressure (larger number of mutations corresponds to smaller fitness, and hence smaller abundance of OTU). The remaining  $(1 - \alpha)N$  OTUs (neutral OTUs) are randomly distributed throughout the sequence space, and they represent the sequences undergoing dynamics under no evolutionary pressure (neutral dynamics).

Each OTU in the model system is associated with an abundance. The abundances of neutral OTUs are randomly sampled from an exponential distribution. (In the Hubbell Neutral Model, the OTU rank abundances are exponentially distributed.) On the other hand, the abundance of niche OTUs exponentially scales with their closeness to the niche:

$$N_i = A \exp(-d_i) \quad [2]$$

where  $N_i$  is the abundance of OTU  $i$  and  $d_i$  is the distance from the OTU to the center of the niche (in sequence space). The results of our metric, the distributions of  $\{E_k\}$  are shown in Fig. 3 for 3 model systems characterized by values of  $\alpha = 0, 0.5$  and  $1$ . We see that the heuristic arguments we described in the previous section and sketched out in Fig. 1(c) are consistent with these model numerical calculations.



**Fig. 3.** The results of our metric, the distributions of  $E$  shown for a fully Niche-like model dataset ( $\alpha = 1$ ), a fully Neutral-like model dataset ( $\alpha = 0$ ) and an intermediate dataset ( $\alpha = 0.5$ ). The results shown are the average of 5000 Monte Carlo simulations for each dataset.

It is instructive to demonstrate the effects of two factors on our metric, in order to highlight some of the mathematical considerations that went into the design of the metric, in particular our use of an extremal measure (the shortest distance aspect of our metric) and the influence of sampled abundance distributions. First we demonstrate the role of extremality introduced by choosing the subset  $\{E\}$ . Instead, if we choose to plot the distribution of  $\{H\}$  we obtain qualitatively the same results for neutral-like models (compare models 1 and 2 in Fig. S9). However, for niche-like models, the peak at zero moves to a nonzero peak which corresponds to the average size of the niche (compare models 5 and 6 in Fig. S9). Thus, the choice of an extremal measure is important in making sure that the endmember distributions (pure niche, pure neutral) are clearly distinct.

Second, we demonstrate what might appear at first to be a rather counter-intuitive fact: the distribution of distances is only weakly dependent on the abundance distribution of the OTUs. If the abundance of an OTU  $k$  is  $N_k$  then we could imagine modifying our procedure by weighting the contribution of  $E_k$  in the distribution  $\{E\}$  by a factor of  $N_k$ . Such a weighting introduces no change whatsoever to the neutral dataset (compare models 2 and 4 in Fig. S9), and no qualitative change in the niche dataset (models 6 and 8 in Fig. S9). Finally, we can also weigh the distribution of  $\{H\}$  in such a way that each distance  $H_{ij}$  between OTUs  $i$  and  $j$  gets weighted by a factor of  $N_i N_j$ . The results are exactly the same as with no weighting for the neutral dataset (compare models 1 and 3 in Fig. S9) and qualitatively the same for the niche dataset (compare models 5 and 7 in Fig. S9).

## Results

We performed a pyrosequencing study of the gastrointestinal (GI) microbiomes of 3 pairs of domesticated vertebrates: 2 swine, 2 cattle and 2 chickens. These pairs of organisms were chosen as pilots for probing specific microbiome issues of relevance to animal science. In particular, we attempted a comparative study looking at the effects of diet on identically cloned swine, and the effects of a microbial challenge on two identically-raised chickens. For the purposes of this paper, these comparisons and the outcomes of the experiments are not of interest: full details of the comparisons and other studies will be published elsewhere. In this study, two genetically identical cloned swine were fed different diets and then their fecal samples were collected for sequencing. Cattle rumen 1 and cattle rumen 2 were rumen fistula sampled at 0 and 8 hours after feeding, respectively (51). Chicken caecum 94 was inoculated with *Campylobacter jejuni* one week prior to caecal sampling. Chicken caecum 1 was kept under the same conditions but without oral gavage of *C. jejuni* (52). See the Methods for details regarding the laboratory protocols. The GI Samples were subjected to deep hypervariable 16S rRNA tag sequencing using a 454 Life Science Genome Sequencer GS FLX (49). Supplementary Table S1 shows the average read length and number of reads obtained for each sample.

Following their acquisition, we aligned the pyrosequenced reads using NAST (53) to the SILVA (54) database. We also aligned the reads using RDP's frontend (55) to the Infernal (56) structural aligner. For each dataset, the NAST+SILVA and RDP+Infernal multiple alignments were merged and hand curated using the methodology and tools described in Sipos *et al.* (48). Short reads and sequences with unknown nucleotides were removed. Spurious "tails" in the multiple alignment, sequences that extend beyond the region of 16S common to all the sequences in the dataset, were also removed. Distance matrices were generated from the multiple alignments, and were then fed to a complete linkage clustering algorithm to generate the OTUs. The careful multiple alignment procedure led to a vast reduction in the number of resulting OTUs in the datasets as previously reported in Sipos *et al.* (48). See table S1 for multiple alignment, species diversity and richness metrics for each of the 6 GI micro-

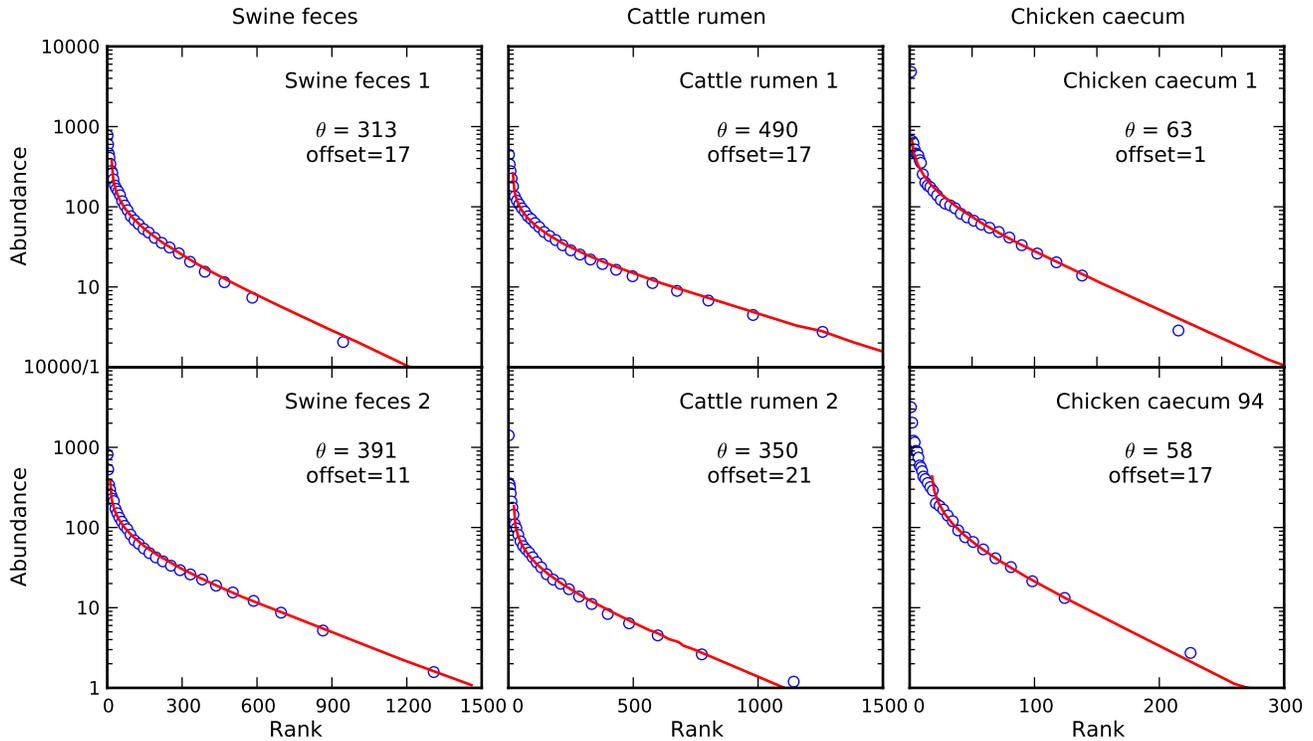
biome samples. Rarefaction curves show how the number of sampled OTUs varies as a function of the number of organisms sampled. Our rarefaction curves are shown in Fig. S2 for each of the 6 datasets.

We plotted the abundances of the OTUs for each of the 6 datasets in our study, and we find a very good agreement with the Neutral Model. These are displayed in rank-abundance form in Fig. 4, and in alternative forms in Fig. S3 and S4. The early ranks (high abundance OTUs) show some systematic deviation from the abundances expected from neutral theory but at face value, these results are consistent with the majority portion (thousands) of the OTUs evolving in the absence of any apparent selection acting on the individual OTUs. Given all the factors that influence the gastrointestinal microbiome (57–62), and the reproducible, thereby seemingly host-selected, microbial abundances (63), it seems counterintuitive that there should be no apparent selection for the vast majority of OTUs in the exponential tail of the rank abundance. However, if we compare taxonomic assignments of microbes across each pair of animals in our study (Fig. S5), we find that there is a correlation between the relative abundances of taxa in members of each animal pair. Namely, we observe that the most abundant taxonomic orders are the same for each animal pair (Clostridiales for swine and chickens, and Pseudomonadales for cattle). This correlation also extends to other taxonomic orders. Hence, our dataset indicates that certain taxa are favored more than others within the GI tract of these 6 vertebrates.

We now attempt to resolve this apparent contradiction, namely that the Neutral Theory fits the rank abundance patterns well, with only 2 fitting parameters, even though the taxonomic data suggests Niche selection. In order to do this, we must turn our attention to other information contained within the pyrosequenced reads. As shown in Fig. 1 the OTUs with their characteristic sequences and associated abundances form patterns within a high-dimensional space. Each read constitutes a point in this space, defined by its nucleotide sequence. One way in which we can attempt to comprehend the structure of this space is through dimensional reduction. We use Principal Component Analysis (PCA) in order to place the OTUs into a 2 dimensional space spanned by the two principal components. We perform a weighted version of PCA (64) where we assign a weight to the OTUs proportional to their abundance. The resulting patterns in the space of two principal components are shown in Fig. S6. Each circle in the figure is an OTU and the circles' size and color indicates the logarithm of the OTU abundance.

As a control, we generate datasets of artificially generated sequences (hereafter referred to as neutral datasets). We generate a neutral dataset for each of the 6 experimental datasets to facilitate a 1-to-1 comparison. Each neutral dataset is constructed in a way such that it has the same number of OTUs and the same OTU abundance distribution as the associated experimental dataset. However, the representative sequence for each OTU is artificially generated and has a randomized sequence, with constraint such that it has the same sequence statistics as the original dataset (probability of observing a nucleotide at a position in the multiple alignment) and column conservation. This ensures that the sequences are randomly distributed along a realistic sub-manifold of sequence space (the subset of 16S sequences that are allowed by secondary structure). We then run the PCA on the neutral datasets (Fig. S7). Comparing Fig. S6 and S7, we notice the following pattern in the experimental GI data: the low-abundance OTUs cluster around the high-abundance OTUs in the dimensionally reduced space. In the neutral datasets, this is not observed, instead the PCA distributes the OTUs approximately uniformly in the dimensionally-reduced space.

We now formulate a heuristic to clearly discriminate between the randomly assembled model sequences and those assembled from a niche-driven process. On a rank-abundance curve, we label the  $k\%$  of the most abundant OTUs as modal OTUs. We label the remaining OTUs as rare OTUs (Fig. 2(a)). Instead of using the whole-dataset rank-abundance curve, one can also use per-order rank-abundance curves if additional resolution is necessary. Once modal and rare



**Fig. 4.** Comparison of rank abundance curves and neutral model fits for the six animal GI microbiomes. Lines indicate fits to the Hubbell's neutral metacommunity model. Parameter  $\theta$  of the model is fit to correspond to the exponential tail in rank abundance. Offset represents the number of high-abundance OTUs that do not fit the neutral model.

OTUs have been assigned, for each rare OTU we compute the distance to the modal OTU that is closest to it. The motivation behind this heuristic is the following. The spread pattern of sequence abundances gives us an indication of whether organisms are evolving neutrally or toward defined niches. In long time behavior, Neutral evolution leads to the expectation that organisms have an equal chance of being anywhere in this space. Niche selection, however, suggests a very biased distribution of organisms. In particular, organisms would be densely clustered about the local optimum for each niche (Fig. 2(b)). These two scenarios lead to very different distributions of distance to nearest niche. If the OTUs are undergoing a niche-driven dynamic, then the rare OTUs will tend to drop off exponentially in abundance around the modal OTUs. If on the other hand, the OTUs have been sampled from a community shaped by neutral evolutionary dynamics, then the rare OTUs' distance to closest modal OTU will be peaked around some non-zero distance that is the average distance between any two OTUs in the dataset (Fig. 2(c)).

We apply the above analysis to the case of gastrointestinal microbiome datasets of the 6 vertebrates. The results are summarized in Fig. 5. In this figure, the blue bars indicate the results of our metric applied to experimental data. The dashed red lines indicate the results of the metric applied to a dataset of sequences that were randomized in the way described above. The results indicate that the GI tracts of the 6 vertebrates largely undergo niche dynamics, with the possible exception of a subpopulation of the chicken GI tracts. The chicken datasets have a small non-zero peak corresponding to the average distance between sequences chosen at random. Our study indicates that the sequences within this peak may be undergoing neutral dynamics. The results that we obtain are robust in that they do not qualitatively depend upon the choice of the cutoff  $k$ . In Fig. S8 we show the metric for  $k = 5\%$  and  $k = 7\%$ . Similarly, the results of the metric on model systems are virtually unchanged when  $k$  is changed between 2% and 10% (Fig. S10) indicating robustness. Whereas our metric is

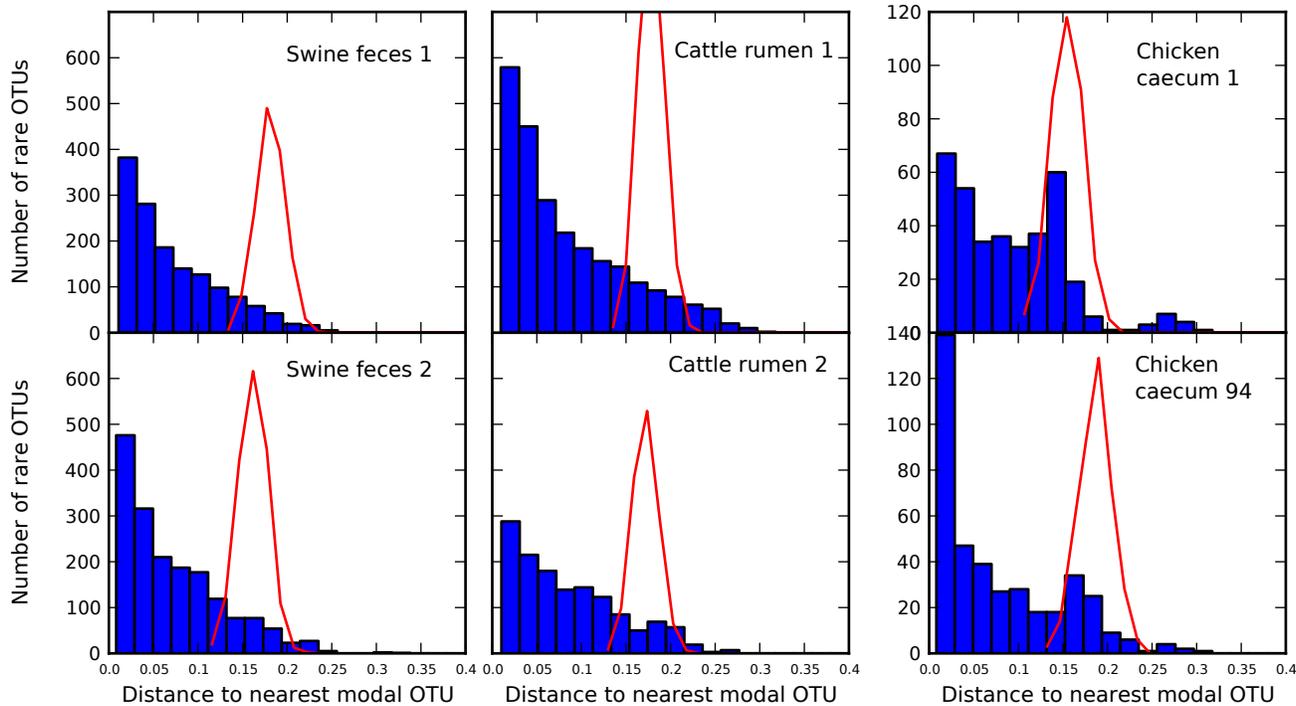
robust in this way, the reader is reminded that phylogenetic resolution is nevertheless important: some niches may appear as a single OTUs at 97% percent sequence identity.

## Discussion

In this work, we set out to construct genomic-based measures of ecosystem diversity and abundance that can provide evidence for process. We focused on understanding the processes that structure microbial communities, because these play functionally important roles in many ecosystems, yet are rich in diversity. Thus, such systems would *a priori* be expected to contain at least sub-populations shaped primarily by stochastic forces. The dual features of high diversity and foundational role functionally in their host ecosystem suggests that microbial communities would not be simple to characterize as either niche or neutral. At the same time theoretical arguments suggest that such high-diversity communities might appear, for fundamental statistical reasons, as neutral.

We succeeded in creating a quantitative metric that fuses abundance and genomic data in order to determine whether an ecological system is dominated by neutral evolution or by niche selection. The key concept was to explore the correlations and associated probability distributions between the most abundant members of the community and the long, low abundance tail members. We showed that the signature of the probability distribution describing the distance in genomic sequence space from each rare OTU to the nearest modal OTU provided a signature of the strength of niche dynamics. We tested this construct on large datasets from 6 animal gastrointestinal tract microbiomes, finding in all cases that the results are inconsistent with neutral assembly. We conclude that niche selection largely dominates within the GI microbiome, despite the fact that the rank abundance patterns are apparently well-modeled by Neutral Theory.

Our results provide firm evidence from an empirical dataset that apparently neutral patterns of diversity and abundance can arise from niche-dominated dynamics, in agreement with earlier theoretical ex-



**Fig. 5.** Histogram of distances of rare OTUs to the nearest modal OTU for each of the 6 gastrointestinal microbiomes with cutoff  $k = 5\%$  (blue bars indicate experimental data). Red dashed lines indicate the results of the metric applied to sequences that were randomized while preserving rank abundance and sequence statistics (see text). Cattle and swine datasets share the same  $y$ -axis.

peptations (2, 5, 19, 27–29). Our results establish definitively that simple ecological measures need to be, and can be, augmented by genomic data in order to provide insight into the processes that structure communities.

## Materials and Methods

**Sample Preparation.** All procedures involving animals were approved by the Institutional Animal Care and Use Committee of the University of Illinois. For each animal, we used two different samples for our test that vary in some aspect such as diet or sampling times. The Duroc sow (2-14; TJ Tabasco) was used as the genomic template for producing cloned animals using somatic cell nuclear transfer. Tabasco was used to produce the CHORI 242 BAC library which was used to generate the full pig genome sequence (65). The clones were born by vaginal delivery and allowed to suckle. They were weaned at 4 weeks of age and continuously housed together. They were not vaccinated or ever in contact with other pigs after weaning. Pigs were fed once daily in the morning and had free access to water. Fecal samples were collected on day 14 (the last day of that feed rotation) of each diet for a total of 4 samples for each animal. Samples were collected from the rectum into a sterile tube and frozen at  $-80^{\circ}\text{C}$  until time of analysis. Bovine rumen samples were collected as previously reported in ref. (51). Chicken caeca were collected as previously reported in ref. (52).

**Sequencing.** Swine and cattle samples were sequenced using PCR product from PCR specific primers flanking the V1-V3 region of bacterial 16S rDNA (66). The forward fusion primers for pyrosequencing included 454 Life Science's A adapter, and barcode A fused to the 5' end of the V1 primer 27F. In chicken the V3 primer 341F was used. In all samples, the reverse fusion primer included 454 Life Science's B adapter (lowercase) fused to 5' end of V3 primer 534R. The fragments in the amplicon libraries were subjected to a single pyrosequence run from the V3 primer end using a 454 Life Science Genome Sequencer GS FLX (Roy J. Carver Biotechnology Center, University of Illinois).

**Rank-abundance, Species-abundance, Preston Plots and Taxa Distributions.** The reads from cattle and swine microbiomes were cleaned up using the method recommended in ref. (67). For the chicken caecum microbiome we

removed all sequences shorter than 100 bp. The ends of all reads were trimmed so that the sequences start and end in the same place in the 16S rRNA consensus structure. All remaining sequences were then aligned using the method described in ref. (48). The OTUs were clustered using complete linkage (68) 3% sequence identity with the denominator 4 from (69) (counting indels as differences). The OTU abundance data for rank-abundance was then binned into a histogram using the method in Adami and Chu (70). Species-abundance and Preston plots were generated following ref. (71). Neutral model curves were generated using the algorithm for the sampling organisms from a neutral meta-community (10). Hubbell's  $\theta$  parameter was fixed to match the exponentially decaying tail of the rank abundance. Offset was chosen by a least-squares method. Taxonomy assignments and comparison of libraries was made with the Library Compare tool (72) at RDP (55).

**PCA Ordination.** In Fig. S6 we show the results of Principal Component Analysis on our OTU data. In performing this calculation, each OTU was associated with a vector of real numbers of dimension  $4L$  where  $L$  is the length of the multiple alignment. The elements of the vectors were calculated in the following way. Each nucleotide within the multiple alignment is represented by a sub-vector of 4 numbers, A is (1, 0, 0, 0), C is (0, 1, 0, 0), G is (0, 0, 1, 0), T is (0, 0, 0, 1), whereas the gap is represented as (0, 0, 0, 0). The vector associated with the OTU is then the arithmetic average of the vectors associated with each sequence within the OTU. We then perform the weighted PCA procedure (64) where we weigh each OTU by its abundance.

**Closest-distance metric.** We used the percent sequence distance metric in Fig. 5. The randomized dataset (red line) was generated in the following way. Each OTU (with its associated abundance) was replaced by a representative randomized sequence. This sequence was generated by selecting each nucleotide from a distribution of probabilities generated from the sequence reads. In this way, the base pair distribution for each position in the multiple alignment of the model dataset is the same as that of the experimental dataset. Furthermore, since the abundances of OTUs are kept, the rank abundance of the model dataset is exactly the same as that of the experimental dataset.

**ACKNOWLEDGMENTS.** We thank Carl Woese for insightful discussions during the preparation of this manuscript, Dennis Schaberg for assistance with the inoculation of the chickens with *C. jejuni* and Stuart Perry for animal care. These studies were supported by the Food Safety Research Response Network, a Co-

ordinated Agricultural Project, funded through the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, under grant number 2005-35212-15287, and the USDA/NRI under grant numbers

2006-35206-16652 and 2007-35212-18046. P.J. acknowledges support by the L.S. Edelhelt Family Biological Physics Fellowship. N.C. acknowledges support from the Institute of Genomic Biology Fellows Program.

1. Tokeshi M (1999) *Species coexistence: ecological and evolutionary perspectives* (Wiley-Blackwell, New York).
2. Chesson P (2000) Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* 31:343–366.
3. Hutchinson GE (1959) Homage to Santa Rosalia, or why are there so many kinds of animals? *Am. Nat.* 93:145–159.
4. Hutchinson GE (1961) The paradox of the plankton. *Am. Nat.* 95:137–145.
5. J Chaveé, H C Muller-Landau SAL (2002) Comparing classical community models: theoretical consequences for patterns of diversity. *Am. Nat.* 159:1–23.
6. Silvertown J (2004) Plant coexistence and the niche. *Trends Ecol. Evol.* 19:605–611.
7. Wright S (2002) Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia* 130:1–14.
8. Caswell H (1976) Community structure: a neutral model analysis. *Ecol. Monogr.* 46:327–354.
9. Bell G (2000) The distribution of abundance in neutral communities. *Am. Nat.* 155:606–617.
10. Hubbell S (2001) *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton).
11. Bell G (2001) Neutral macroecology. *Science* 293:2413–2418.
12. Chave J (2004) Neutral theory and community ecology. *Ecol. Lett.* 7:241–253.
13. Rosindell J, Hubbell S, Etienne R (2011) The Unified Neutral Theory of Biodiversity and Biogeography at Age Ten. *Trends Ecol. Evol.* 26:340–348.
14. Munepeeraikul R, et al. (2008) Neutral metacommunity models predict fish diversity patterns in Mississippi–Missouri basin. *Nature* 453:220–222.
15. Woodcock S, et al. (2007) Neutral assembly of bacterial communities. *FEMS Microbiol. Ecol.* 62:171–180.
16. McGill B (2003) A test of the unified neutral theory of biodiversity. *Nature* 422:881–885.
17. McGill B, Maurer B, Weiser M (2006) Empirical evaluation of neutral theory. *Ecology* 87:1411–1423.
18. Hubbell S (2005) Neutral theory in community ecology and the hypothesis of functional equivalence. *Funct. Ecol.* 19:166–172.
19. Leibold M, McPeck M (2006) Coexistence of the niche and neutral perspectives in community ecology. *Ecology* 87:1399–1410.
20. Purves D, Turnbull L (2010) Different but equal: the implausible assumption at the heart of neutral theory. *J. Anim. Ecol.* 79:1215–1225.
21. Ricklefs R (2006) The unified neutral theory of biodiversity: Do the numbers add up? *Ecology* 87:1424–1431.
22. Etienne R, Alonso D, McKane A (2007) The zero-sum assumption in neutral biodiversity theory. *J. Theor. Biol.* 248:522–536.
23. Allouche O, Kadmon R (2009) A general framework for neutral models of community dynamics. *Ecol. Lett.* 12:1287–1297.
24. Adler PB, Rislambars JH, Levine JM (2007) A niche for neutrality. *Ecol. Lett.* 10:95–104.
25. Adler P, Ellner S, Levine J (2010) Coexistence of perennial plants: an embarrassment of niches. *Ecol. Lett.* 13:1019–1029.
26. Volkov I, Banavar J, Hubbell S, Maritan A (2009) Inferring species interactions in tropical forests. *Proc. Natl. Acad. Sci. USA* 106:13854–13859.
27. Purves D, Pacala S, Burslem D, Pinard M, Hartley S (2005) in *Biotic interactions in the tropics: their role in the maintenance of species diversity*, eds Burslem DF, Pinard MA, Hartley SE (Cambridge University Press, Cambridge), pp 107–138.
28. Chisholm R, Pacala S (2010) Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proc. Natl. Acad. Sci. USA* 107:15821–15825.
29. Gravel D, Canham C, Beaudet M, Messier C (2006) Reconciling niche and neutrality: the continuum hypothesis. *Ecol. Lett.* 9:399–409.
30. Tilman D (2004) Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. *Proc. Natl. Acad. Sci. USA* 101:10854–10861.
31. Cadotte M (2007) Concurrent niche and neutral processes in the competition–colonization model of species coexistence. *Proc. R. Soc. B* 274:2739–2744.
32. Zillio T, Condit R (2007) The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. *Oikos* 116:931–940.
33. Loreau M, de Mazancourt C (2008) Species Synchrony and Its Drivers: Neutral and Nonneutral Community Dynamics in Fluctuating Environments. *Amer. Nat.* 172:48–66.
34. Doncaster C, Cornell S (2009) Ecological Equivalence: A Realistic Assumption for Niche Theory as a Testable Alternative to Neutral Theory. *PLoS ONE* 4:e7460.
35. Haegeman B, Loreau M (2011) A mathematical synthesis of niche and neutral theories in community ecology. *J. Theor. Biol.* 269:150–165.
36. Dumbrell A, Nelson M, Helgason T, Dytham C, Fitter A (2009) Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME J.* 4:337–345.
37. Zhang Q, Buckling A, Godfray H (2009) Quantifying the relative importance of niches and neutrality for coexistence in a model microbial system. *Funct. Ecol.* 23:1139–1147.
38. Langenheder S, Székely AJ (2011) Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *ISME J.* 5:1086–1094.
39. Ayarza JM, Erijman L (2011) Balance of neutral and deterministic components in the dynamics of activated sludge floc assembly. *Microb. Ecol.* 61:486–495.
40. Ofiteru ID, et al. (2010) Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci. USA* 107:15345–15350.
41. Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T (2011) Bacterial community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci. USA* 108:14288–14293.
42. Horner-Devine MC, et al. (2007) A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* 88:1345–1353.
43. Emerson B, Gillespie R (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol. Evol.* 23:619–630.
44. Kelly C, Bowler M, Pybus O, Harvey P (2008) Phylogeny, niches, and relative abundance in natural communities. *Ecology* 89:962–970.
45. Cavender-Bares J, Kozak K, Fine P, Kembel S (2009) The merging of community ecology and phylogenetic biology. *Ecol. Lett.* 12:693–715.
46. Kembel S, et al. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26:1463–1464.
47. Cadotte M, et al. (2010) Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. *Ecol. Lett.* 13:96–105.
48. Sipos M, et al. (2010) Robust Computational Analysis of rRNA Hypervariable Tag Datasets. *PLoS ONE* 5:e15220.
49. Huse SM, et al. (2008) Exploring microbial diversity and taxonomy using ssu rna hypervariable tag sequencing. *PLoS Genet.* 4:e1000255.
50. Badger JH, Ng PC, Venter JC (2011) in *Metagenomics of the Human Body*, ed Nelson KE (Springer, New York), pp 1–14.
51. Brulc JM, et al. (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci. USA* 106:1948–1953.
52. Qu A, et al. (2008) Comparative Metagenomics Reveals Host Specific Metaviruses and Horizontal Gene Transfer Elements in the Chicken Cecum Microbiome. *PLoS ONE* 3:e2945.
53. DeSantis TZ, et al. (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res.* 34:394–399.
54. Pruesse E, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35:7188–7196.
55. Cole JR, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37:D141–D145.
56. Nawrocki EP, Kolbe DL, Eddy SR (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25:1335–1337.
57. Bäckhed F, Ley R, Sonnenburg J, Peterson D, Gordon J (2005) Host-bacterial mutualism in the human intestine. *Science* 307:1915–1920.
58. Dethlefsen L, McFall-Ngai M, Relman D (2007) An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* 449:811–818.
59. Turnbaugh P, et al. (2007) The human microbiome project. *Nature* 449:804–810.
60. Turnbaugh P, et al. (2008) A core gut microbiome in obese and lean twins. *Nature* 457:480–484.
61. Li M, et al. (2008) Symbiotic gut microbes modulate human metabolic phenotypes. *Proc. Natl. Acad. Sci. USA* 105:2117–2122.
62. Slack E, et al. (2009) Innate and Adaptive Immunity Cooperate Flexibly to Maintain Host-Microbiota Mutualism. *Science* 325:617–620.
63. Antonopoulos D, et al. (2009) Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. *Infect. Immun.* 77:2367–2375.
64. Kriegel HP, Kröger P, Schubert E, Zimek A (2008) in *Scientific and Statistical Database Management*, Lecture Notes in Computer Science, eds Ludäscher B, Mamoulis N (Springer, Heidelberg) Vol. 5069, pp 418–435.
65. Humphray S, et al. (2007) A high utility integrated map of the pig genome. *Genome Biol.* 8:R139.
66. Muyzer G, Dewaal E, Uitterlinden A (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59:695–700.
67. Kunin V, Engelbrekton A, Ochman H, Hugenholtz P (2009) Wrinkles in the rare biosphere: pyrosequencing errors lead to artificial inflation of diversity estimates. *Environ. Microbiol.* pp 118–123.
68. Schloss P, et al. (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* 75:7537–7541.
69. May ACW (2004) Percent sequence identity; the need to be explicit. *Structure* 12:737–738.
70. Adami C, Chu J (2002) Critical and near-critical branching processes. *Phys. Rev. E* 66:011907.
71. Gray J, Bjorgesaeter A, Ugland K (2006) On plotting species abundance distributions. *J. Anim. Ecol.* 75:752–756.
72. Wang Q, Garrity G, Tiedje J, Cole J (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73:5261–5267.

# **Supplemental material for “Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes”**

Patricio Jeraldo<sup>1,2\*</sup>, Maksim Sipos<sup>1,2\*</sup>, Nicholas Chia<sup>1,2</sup>, Jennifer M. Brulc<sup>1,3</sup>, A. Singh Dhillon<sup>4</sup>, Michael E. Konkel<sup>5</sup>, Charles L. Larson<sup>5</sup>, Karen E. Nelson<sup>6</sup>, Ani Qu<sup>1,3,7</sup>, Lawrence B. Schook<sup>1,3</sup>, Fang Yang<sup>1,3</sup>, Bryan A. White<sup>1,3</sup> & Nigel Goldenfeld<sup>1,2</sup>

<sup>1</sup>*Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, IL 61801, USA*

<sup>2</sup>*Loomis Laboratory of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801, USA*

<sup>3</sup>*Department of Animal Sciences, 1207 West Gregory Drive, Urbana, IL 61801, USA*

<sup>4</sup>*Washington State University Avian Health and Food Safety Laboratory, Washington State University, Puyallup, WA 98371, USA*

<sup>5</sup>*School of Molecular Biosciences, Washington State University, Pullman, WA 99164, USA*

<sup>6</sup>*The J. Craig Venter Institute, 9712 Medical Center Drive, Rockville, MD 20850, USA*

<sup>7</sup>*Environmental and Occupational Health Sciences Institute, Piscataway, NJ 08854, USA*

	<b># reads</b>	<b>Unique reads</b>	<b>Average length</b>	<b>Aligned width</b>	<b># OTUs at 3%</b>	<b>Simpson diversity</b>	<b>Shannon diversity</b>	<b>Jackknife richness</b>	<b>ACE richness</b>	<b>Chao1 richness</b>
Swine feces 1	33283	14122	165.0	420	1509	0.0070 ±0.0003	5.8 ±0.02	2000 ±260	1472 ±55	1540 ±150
Swine feces 2	36254	16198	175.3	418	1856	0.0068 ±0.0003	5.9 ±0.02	2300 ±300	1633 ±53	1720 ±150
Cattle rumen 1	31201	18264	180.7	471	2580	0.0044 ±0.0002	6.3 ±0.02	3300 ±260	3070 ±88	2640 ±190
Cattle rumen 2	19642	10074	183.6	385	1509	0.0110 ±0.0006	5.9 ±0.03	2070 ±110	1818 ±62	1830 ±130
Chicken caecum 1	17585	2151	136.5	310	396	0.084 ±0.003	4 ±0.03	770 ±120	655 ±75	620 ±150
Chicken caecum 94	21646	2223	138.9	332	354	0.046 ±0.001	3.9 ±0.02	560 ±90	426 ±57	460 ±100

Table S1: Summary statistics of our six datasets.

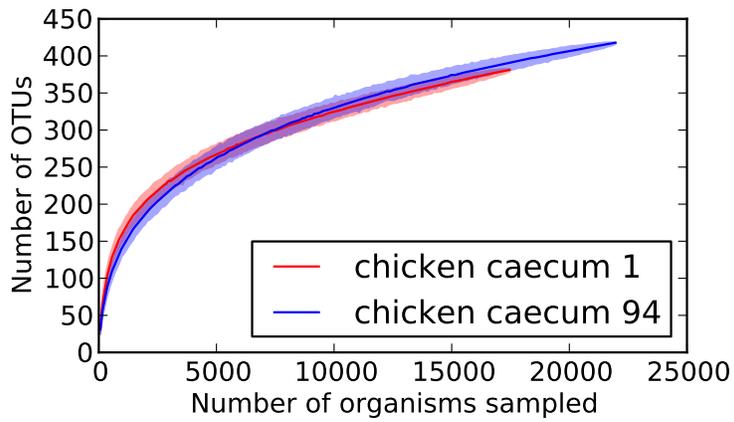
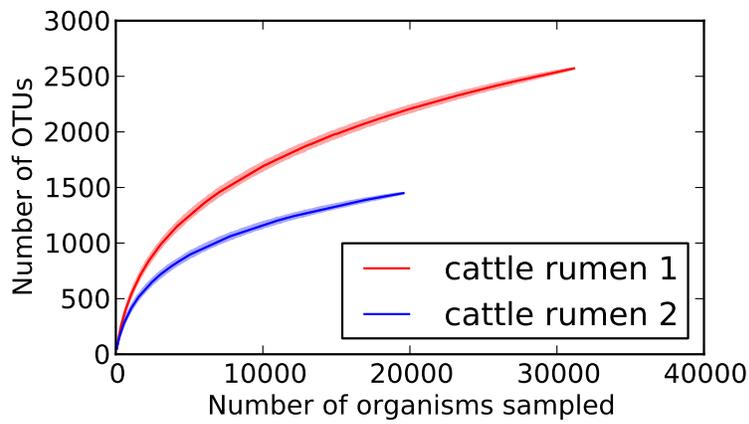
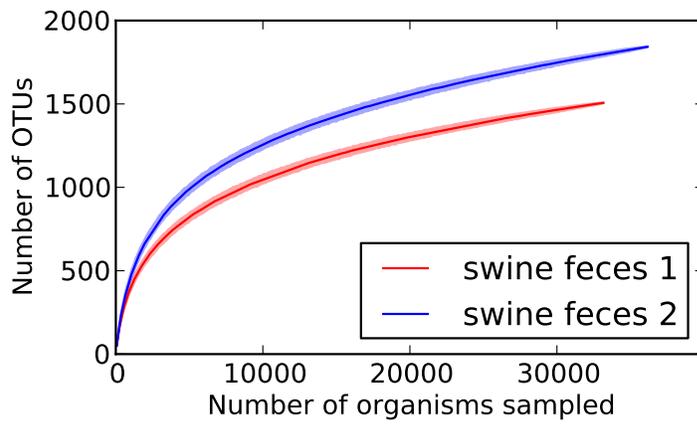


Figure S2: Rarefaction curves for the 6 vertebrate GI microbiomes. Solid line represents the median number of OTUs (100 resamplings) whereas the shaded area represents the 95% confidence interval.

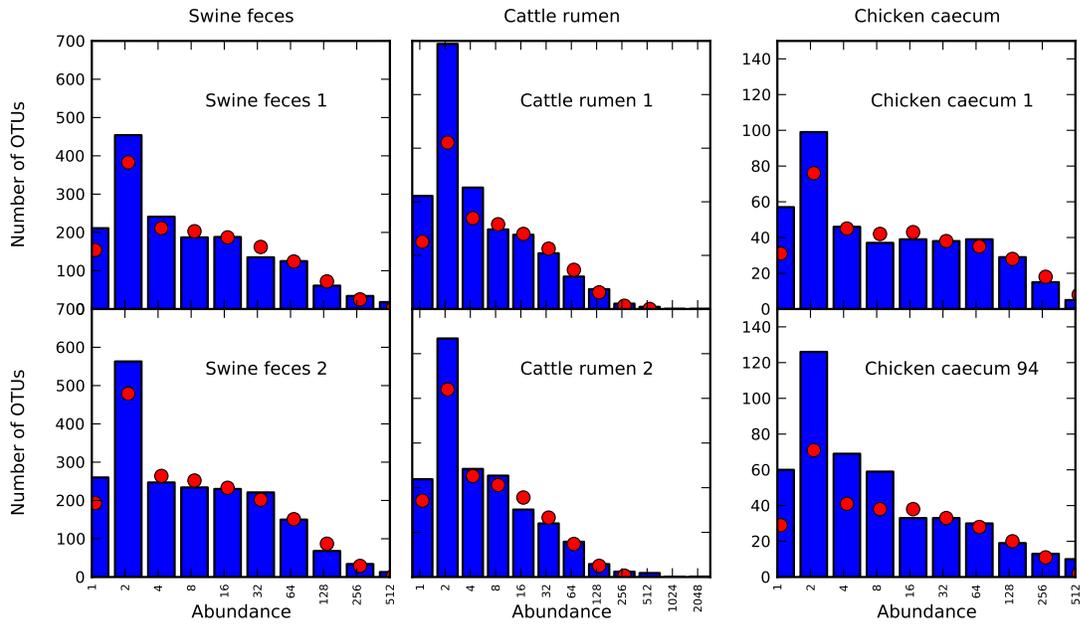


Figure S3: Preston plot for swine feces, cattle rumen and chicken caeca samples. In a Preston plot, the height of the bar indicates the number of species observed with abundance 1, 1-2, 2-4, 4-8, etc. Note that in all 6 datasets most OTUs are singletons. In this plot, 1-2 bars are highest because of an artifact. Traditionally, in a Preston plot, the OTUs with borderline abundances split evenly between two neighboring bins.

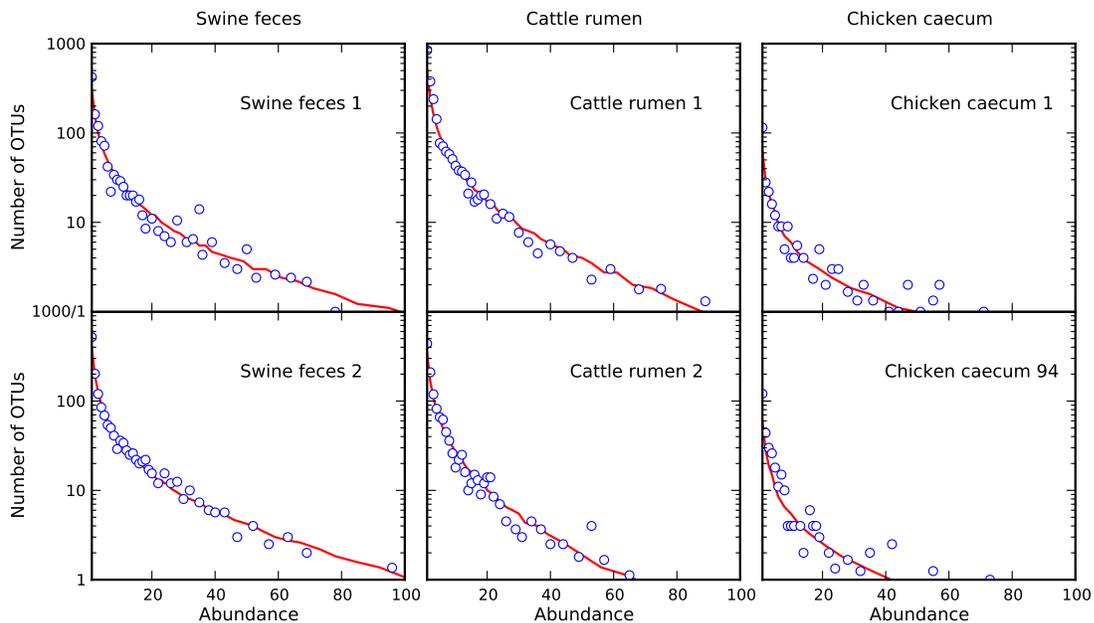


Figure S4: Species abundance distribution for swine feces, cattle rumen and chicken caeca. The species abundance distribution indicates the number of OTUs collected for each abundance.

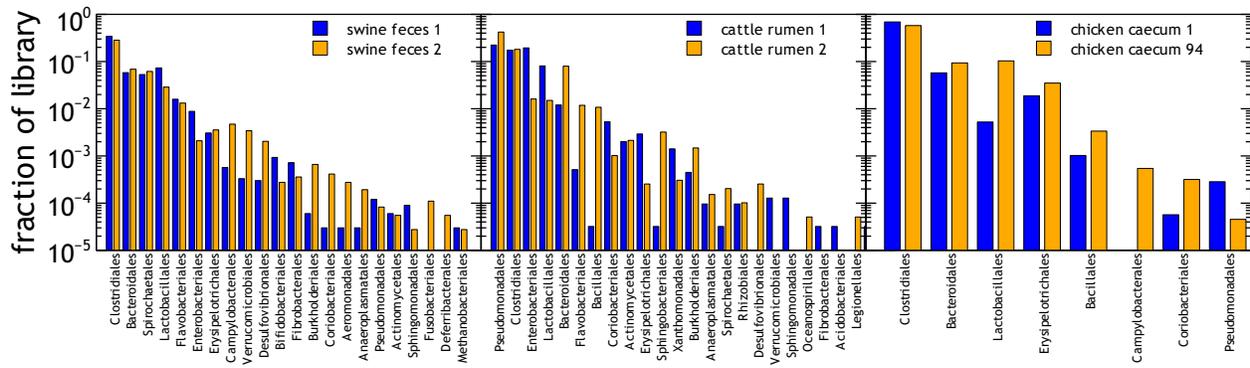


Figure S5: Taxa Comparisons. Taxonomic assignments at order level for all libraries, at 80% confidence threshold, sorted by combined abundance. Though there appear to be no differences in the form of the rank-abundance curves, we see differences in the taxonomic distributions here as the result of changes in diet or challenges to the microbial ecosystem.

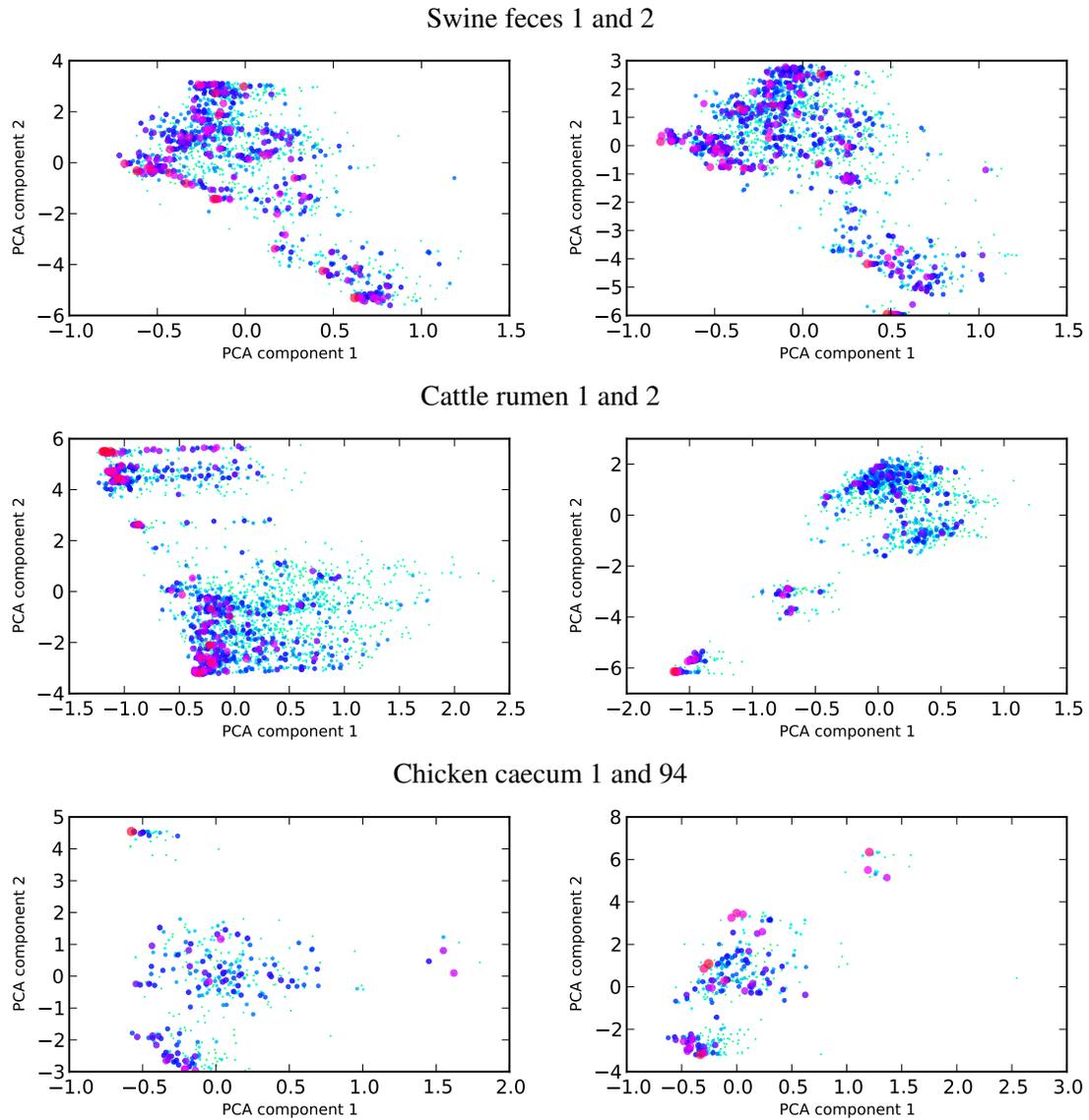


Figure S6: Weighted PCA ordination applied to the 6 experimental datasets. See the main text for details on how weighted PCA was performed. Each circle in this Figure represents an OTU and its size and color indicates the logarithm of OTU abundance.

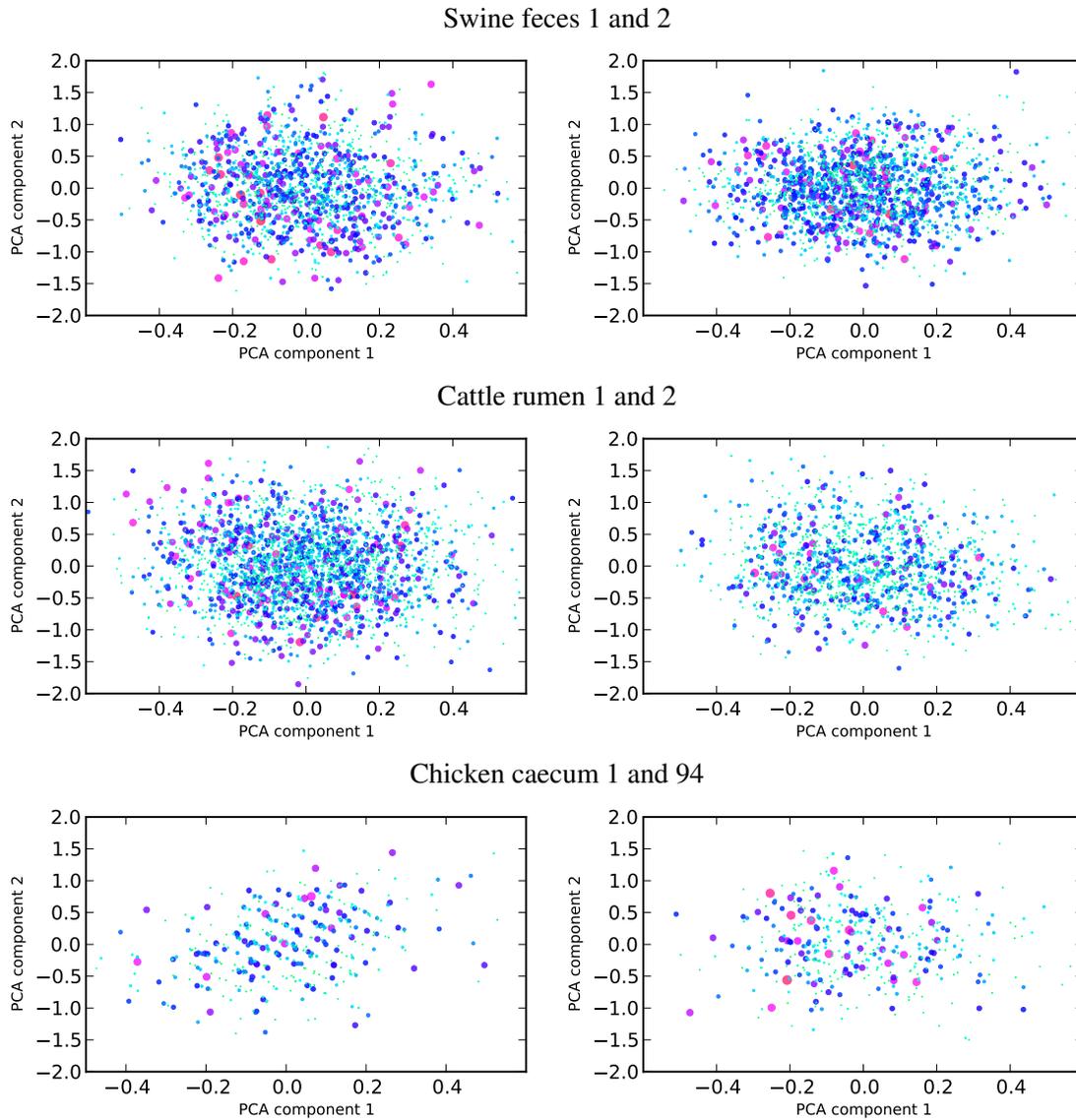
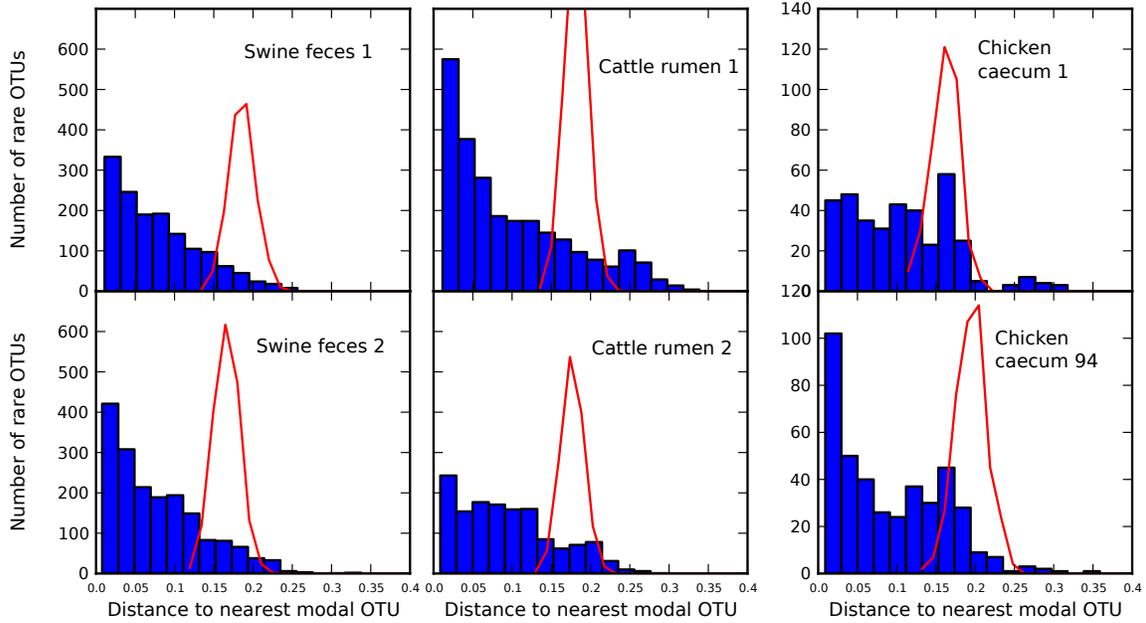


Figure S7: Weighted PCA ordination applied to the randomized datasets. Compare with Fig. S5. See the main text for details on how the randomized datasets were generated, and how weighted PCA was performed. Each circle in this Figure represents an OTU and its size and color indicates the logarithm of OTU abundance.

$k = 3\%$



$k = 7\%$

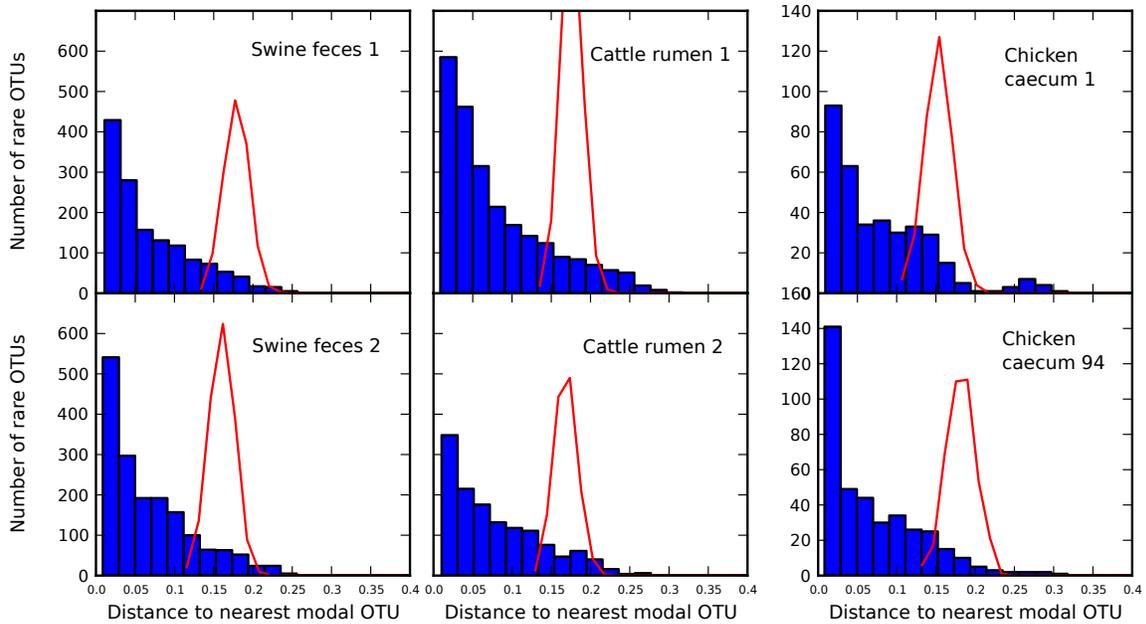


Figure S8: Histogram of distances of rare OTUs to the nearest modal OTU for each of the 6 gastrointestinal microbiomes with cutoffs  $k = 3\%$  and  $k = 7\%$ . Red dashed lines indicate the results of the metric applied to sequences that were randomized while preserving rank abundance and sequence statistics (see main text).

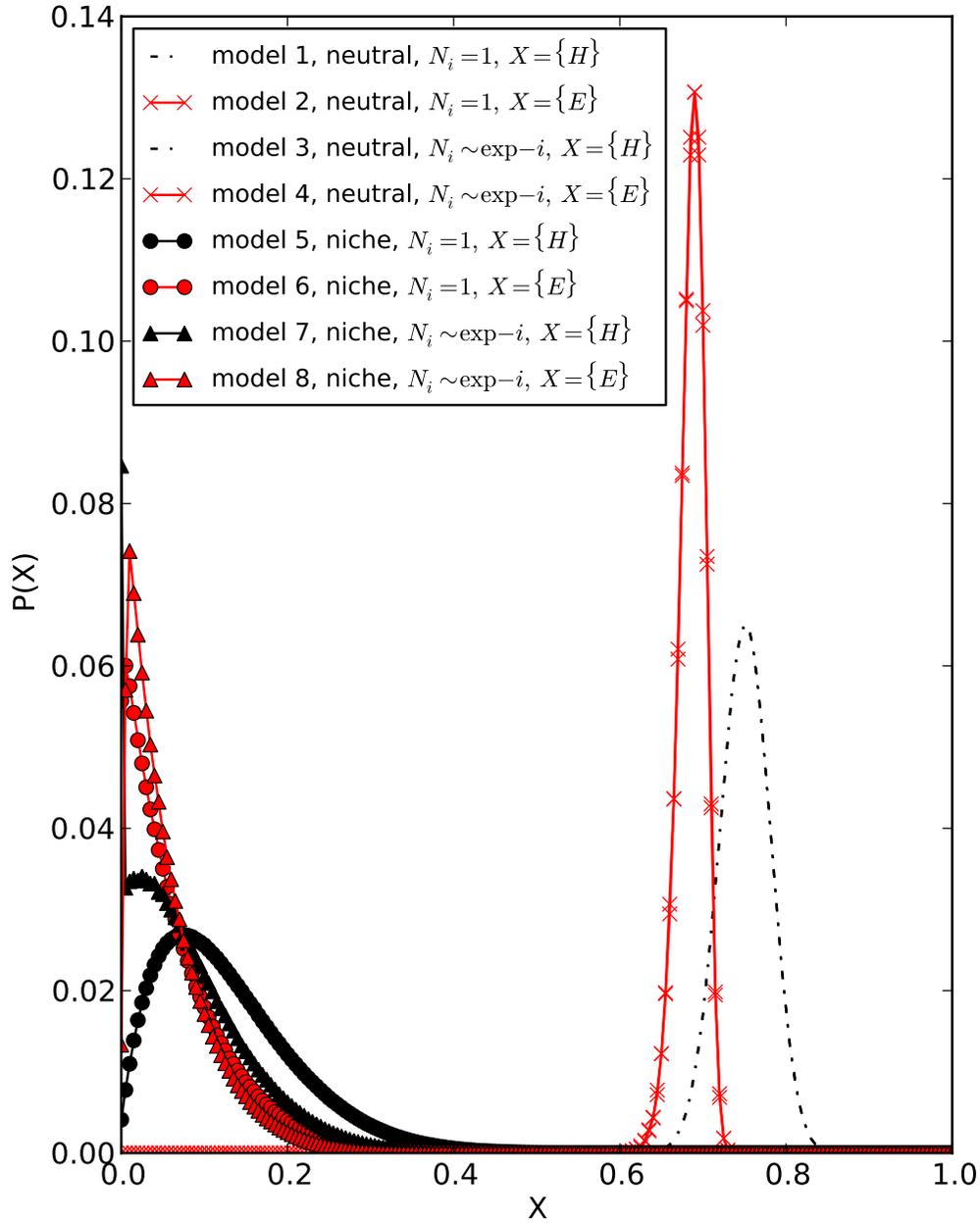


Figure S9: Explicit numerical calculations of our metric on 8 model systems. In these systems, we study the difference between the effects of the metric on neutral (models 1-4) and niche model systems (models 5-8). We also study the effect of choosing the closest distance (even-numbered models) compared to considering all distances (odd-numbered models). Finally, we consider the weighted models (3-4 and 7-8) versus the unweighted ones (1-2 and 5-6).

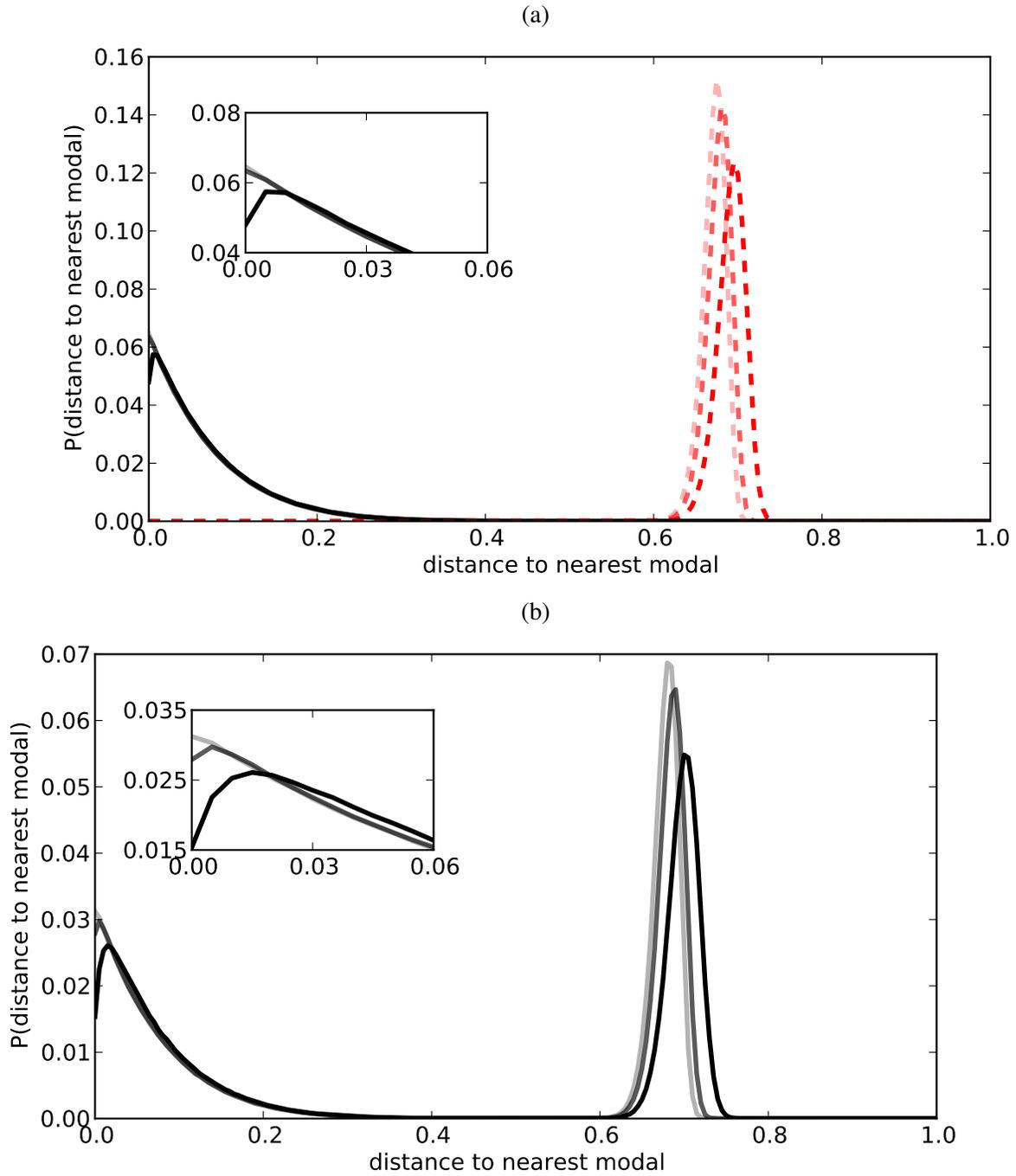


Figure S10: Measuring the effect of the choice of  $k$  on our metric. Darkest lines indicate  $k = 2\%$ , medium lines indicate  $k = 6\%$  and lightest lines indicate  $k = 10\%$ . (a)  $\alpha = 0.0$  model (red dashed lines) and  $\alpha = 1.0$  model (black solid lines). (b)  $\alpha = 0.5$  model (black solid lines).