FEMS Microbiology Ecology, 2022, 98, 1–12
DOI: 10.1093/femsec/fiac021
Advance access publication date: 22 February 2022
Research Article

Microbiome–host-phylogeny relationships in animal gastrointestinal tract microbiomes

Zhanshan (Sam) Ma [1,3,*], Wendy Li [1] and Peng Shi [2,3,*]

[1]Computational Biology and Medical Ecology Lab, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650201, China
[2]Evolutionary and Functional Genomics Lab, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650201, China
[3]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, 650201, China
*Corresponding author: Kunming Institute of Zoology, the Chinese Academy of Sciences No.17 Longxin Road, Kunming, Yunnan, 650201, P.R.China. Tel: +86 871 65130513; E-mail: ma@vandals.uidaho.edu; ship@mail.kiz.ac.cn
One sentence summary: Based on 4903 gastrointestinal-microbiome samples of 318 animal species covering six major invertebrate and all vertebrate classes, we found that power-law patterns seem to govern microbiome-diversity and host-phylogeny–diet relationships.
Editor: Cindy Nakatsu

Abstract

Among the factors influencing the animal gastrointestinal tract microbiome (AGM) diversity, diet and phylogeny have been extensively studied. However, what made the studies particularly challenging is that diet characteristics *per se* are product of evolution, and hence totally disentangling both effects is unrealistic, likely explaining the lack of consensus in existing literatures. To further explore microbial diversity and host-phylogeny–diet relationships, we performed a big-data meta-analysis with 4903 16S rRNA AGM samples from 318 animal species covering all six vertebrate and four major invertebrate classes. We discovered that the relationship between AGM-diversity and phylogenetic timeline (PT) follows a power-law or log-linear model, including diet specific power-law relationships. The log-linear nature predicts a generally rising trend of AGM diversity along the evolutionary tree starting from the root, which explains the observation why *Mammalia* exhibited the highest AGM-diversity. The power-law property suggests that a handful of taxa carry disproportionally large weights to the evolution of diversity patterns than the majority of taxa, which explains why the species richness of *Insecta* was significant different than those from the other nine classes. Finally, we hypothesize that the diversity—PT power-law relationship explains why species-abundance distributions generally follow highly skewed probability distributions.

Keywords: animal gastrointestinal tract microbiome (AGM), animal–microbe co-evolution, diversity–phylogenetic timeline (DPT) log-linear model, diversity–phylogenetic timeline (DPT) power-law relationship, Hill numbers

Introduction

The co-evolution between animal microbiomes and their hosts has become a focus in animal microbiome research in recent years. The *hologenome* theory of evolution (Rosenberg et al. 2009, Theis et al. 2016, Rosenberg and Zilber-Rosenberg 2018) considers the individual animal or plant as a community or holobiont consisting of the host and all of its symbiotic microbes. The hologenome, a collective genome carried by the holobiont, which includes microbial genomes, can be inherited between generations with reasonable fidelity. The variations in the hologenome are subject to natural selection and genetic drifts (Rosenberg et al. 2009, Rosenberg and Zilber-Rosenberg 2018). For example, it has been revealed that gut microbiome may play an important role in speciation (Brucker and Bordenstein 2013). Furthermore, changes in either the host genome or the microbiome (such as new acquisitions of microbes, horizontal gene transfers, and changes in microbial species abundances within hosts) can lead to genetic variation in the hologenome. The changes in microbial species abundances within and across hosts can best be characterized by microbiome diversity patterns, the very topic of the present article.

One of the most important co-evolutionary forces between the animal gastrointestinal tract microbiome (AGM) and its host is likely the trophic or feeding relationship because gut microbes are critical for host digestion of their foods. The three major diet types of animals, i.e. herbivore, carnivore, and omnivore, occupy different levels in ecological communities (ecosystems). They form the backbone of the community (ecosystem) both structurally and functionally, and their symbiotic microbes may modulate both co-operation and competition within the holobiont and with other holobionts. The biodiversity, as a fundamental property of the AGM, together with its relationships with host-phylogeny and diet types, are frequently the first attempt to tackle such complex co-evolutionary interactions. It is important to note that the concepts (or principles) of co-evolution, holobionts, and hologenomes do not necessarily 'map' to each other completely. For example, some microbes (and the genes they carry) may be 'passengers', rather than 'drivers', and 'passengers' may not be involved in co-adaptation or co-evolution processes.

Among the factors influencing the AGM-diversity, diet and phylogeny have been extensively studied (Ley et al. 2008, Degnan et al. 2012, Delsuc et al. 2014, Carmody et al. 2015, Vital et al. 2015; Moeller et al. 2014, 2016, Groussin et al. 2017, Martinson et al. 2017, Gaulke et al. 2018, Mazel et al. 2018, Sherrill-Mix et al. 2018, Amato et al. 2019, Gomez et al. 2019, Youngblut et al. 2019). However, what made the studies particularly challenging is that diet characteristics *per se* are product of evolution, and hence totally disentangling

Received: August 17, 2021. **Revised:** February 8, 2022. **Accepted:** February 18, 2022
© The Author(s) 2022. Published by Oxford University Press on behalf of FEMS. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

both effects may be unrealistic, explaining the lack of consensus in existing literature. Some studies suggested no association between host-phylogeny and the gut microbiome (Degnan et al. 2012, Carmody et al. 2015, Martinson et al. 2017), but many others have confirmed the existence of such a relationship (Moeller et al. 2014, 2016, Gaulke et al. 2018, Sherrill-Mix et al. 2018, Gomez et al. 2019, Youngblut et al. 2019, Song et al. 2020, Trevelline et al. 2020). Since diet characteristics *per se* are product of evolution, studies on the influences of diet type and phylogeny on gut microbiomes are frequently integrated (Ley et al. 2008, Muegge et al. 2011, Delsuc et al. 2014, Vital et al. 2015, Amato et al. 2019); however, it can be extremely challenging to disentangle the interdependency between both phylogeny and diet types. For example, one of the pioneering studies conducted by Ley et al. (2008), which analyzed the gut microbiome of 59 Mammalia species, found that the structure of the gut microbiome was related to both host-phylogeny and diet, and herbivores exhibited the highest diversity, followed by omnivores and carnivores. The same team later illustrated that the processes of microbiome adaptation (evolution) to diet are similar across various mammalian lineages (Muegge et al. 2011).

Some studies also suggested that host diet and evolutionary history both significantly influence the structure of gut microbiome, but aspects of effects could be different. Groussin et al. (2017) showed that phylogeny and diet types could differentially influence bacterial lineages in gut; diet may be associated with the ancient and large bacterial lineages, but phylogeny is mostly associated with recently diverged bacterial lineages. Besides diet and evolutionary history, studies also demonstrated that geography of hosts could affect the composition of the gut microbiome, indicating that the physical distance between hosts can limit the dispersal of the gut microbiome (Moeller et al. 2017). Habitats are also considered to be important determinants of the microbial structure, since terrestrial and aquatic mammals harbor significantly different microbiomes.

Here, we leveraged AGM datasets from over 300 animal species spanning 10 major animal classes (including all six vertebrate classes) to conduct the largest meta-analysis of host-phylogeny and diet, and AGM-diversity. Specifically, the 4903 samples were collected from 318 animal species, which are distributed geographically over all five major continents. Note that we obtained the datasets from 91 existing literatures (see Table S7 (Supporting Information) for the data sources and other important sample information). Using this extensive dataset, we construct quantitative models between animal (host) phylogenetic timeline (PT) and AGM diversities for major host diet types. PT is an alternative concept (metric) to familiar phylogenetic distance (PD). The difference between them is that the PD measures the divergence time for a pair of taxa, while the PT, also known as evolutionary timeline (ET), can be considered as 'age' of a taxon with ancient taxa having larger PT values and modern taxa having smaller PT values (Kumar et al. 2017, Wolfe et al. 2019, Matschiner et al. 2020).

The significant expansion of AGM datasets and the inclusion of AGM-diversity–phylogenetic-timeline (DPT) relationships may significantly expand our understanding on the AGM diversity patterns, and resolve a few significant issues surrounding the AGM research. For example, in pre-experiments, we found that the DPT follows a power-law relationship. If the DPT power-law relationship can pass rigorous statistical tests with the AGM datasets, we expect that it may explain or answer rather important questions in evolutionary ecology. For example, are different animal taxa equally capable in hosting equally diverse microbiomes? Do ancient animal species host more or less diverse microbiomes than modern animal lineages? Why is the species abundance distribution (SAD), whether it is host or microbial species, usually highly skewed (e.g. log-normal, log-series, and power-law distributions)? Of course, we realize that extreme caution should be taken to make causal inferences from simple correlation (regression) relationships. To the best of our knowledge, our focus on quantitative DPT relationship has not been approached in the existing literature. From a broader perspective, our study may also shed light on the co-evolution between microbiomes and their animal hosts.

Methodologically, besides building DPT models, we adopt two recent important advances in biodiversity research. First, we utilized what is considered as the most appropriate metric for measuring alpha-diversity, i.e. the Hill numbers (Chao et al. 2012, 2014a,b), to avoid potential inconsistencies from choosing various diversity metrics such as species richness, Shannon entropy and/or Simpson's index. The Hill numbers possess a few key advantages over the traditional diversity indexes. For example, Hill numbers, which are based on Renyi entropy, standardize the diversity as 'species equivalents' with different levels of *commonness* (*rarity*) of species abundances and stratify the diversity comparisons to corresponding levels (known as diversity orders $q = 0, 1, 2\ldots$). This property greatly simplifies the comparisons of diversity and potential diversity estimates and makes our inferences much more robust against major error sources that might be present in a large-scale meta-analysis such as ours. Second, we apply the rarefaction approach for estimating the Hill numbers (diversity) in order to deal with the so-termed 'sampling problem'—potential inaccuracies in diversity estimations from using community samples obtained with different levels of 'sampling efforts' (e.g. different numbers of 16S-rRNA sequencing reads; Chao et al. 2014b, Ma 2017, Ma and Li 2018, Ma et al. 2019).

Materials and methods

The datasets of AGM

The 16S rRNA sequencing reads datasets of more than 6900 samples of AGM were collected from 108 published studies, covering five phyla and 19 classes (see Table 1 below and Table S7 (Supporting Information) in the Online Supplementary Information (OSI)). The samples collected from other animal body sites rather than from gastrointestinal tract (including stool) were removed, and the samples whose hosts were treated with antibiotics were also removed. To balance sample sizes and the distribution across taxa, the animal classes with less than 10 samples were excluded from the analysis (i.e. a total of nine classes including *Gastropoda*, *Bivalvia*, *Polychaeta*, *Clitellata*, *Diplopoda*, *Merostomata*, *Asteroidea*, *Echinoidea*, and *Holothuroidea* were excluded).

To fully take advantages of rarefaction approach (Chao et al. 2014b) in dealing with the so-termed 'sampling problem'—the diversity estimation can be highly sensitive to the sample sizes, the samples with less than 2000 16S rRNA reads were removed. After the above-described data quality control steps, as shown in Table 1, 4903 samples collected from 91 studies remained, covering three primary phyla (*Nematoda*, *Arthropoda*, and *Chordates*), 10 classes (*Chromadorea*, *Arachnida*, *Malacostraca*, *Insecta*, *Chondrichthyes*, *Actinopteri*, *Amphibia*, *Sauropsida Aves*, and *Mammalia*), and 318 animal species. That is, the selected datasets cover all six classes of vertebrates and two most important phyla (*Nematoda* and *Arthropoda*) of invertebrates, and therefore, are rather representative for the animal kingdom.

Among the 4903 AGM samples, 1421, 1229, and 1473 samples were collected from carnivore, herbivore, and omnivore groups, respectively. The remaining 320 samples were collected from an-

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

Table 1. A brief summary on the AGM datasets collected from the 318 animal species spanning 10 animal classes (see Table S7 (Supporting Information) for detailed information).

		Animal taxa				Number of samples from each diet type				Total number of samples
Animal phylum	**Animal class**	**Orders**	**Families**	**Genera**	**Species**	**Carnivore**	**Herbivore**	**Omnivore**	**Other**	
Nematoda	*Chromadorea*	1	2	2	4	0	0	0	215	215
Arthropoda	*Arachnida*	1	5	8	8	20	0	0	25	45
	Malacostraca	3	6	8	7	0	14	12	4	30
	Insecta	10	33	53	76	4	755	168	52	979
Chordates	*Chondrichthyes*	2	2	4	5	32	0	0	0	32
	Actinopteri	9	12	38	51	911	112	248	0	1271
	Amphibia	1	2	3	3	13	0	8	0	21
	Sauropsida	2	6	9	12	88	88	132	0	308
	Aves	15	20	27	29	13	88	395	7	503
	Mammalia	15	54	109	123	393	564	542	0	1499
Total	Frequency = 10	59	142	261	318	1474	1621	1505	303	4903

imals that do not belong to any of the previous three diet types, and were lumped together as the 'Other' group. A brief summary on the datasets is listed in Table 1, and more detailed information is listed in Table S7 (Supporting Information), in which all study-accession numbers for data access were presented. From the 16S rRNA raw sequencing reads, we utilized QIIME-2 (Version 2018.6.0: Bolyen et al. 2018) software package to compute the operational taxonomic unit (OTU) tables at the 97% (species) similarity level. Phylogenies were analyzed and visualized with the ape (Paradis et al. 2004) and ggtree (Yu et al. 2017) R-packages. The phylogenetic tree and phylogeny information were obtained from http://timetree.org (Kumar et al. 2017). This website can compute two types of phylogenetic times: the 'divergence time for a pair of taxa' and an 'evolutionary timeline for a taxon'. The former is known as PD in literature, but we use the latter—ET or PT—as proxy of phylogenetic history of a taxon. One may think it as 'age' of a taxon with ancient taxa having larger PT values and modern taxa having smaller PT values. The PT better fits to our research objectives because it is associated with each taxon, for which we can compute the microbiome biodiversity it hosts.

The Hill numbers as diversity measures

The OTU tables are equivalent to SAD datasets in macrobial ecology of plants and animals, and this conceptual transformation will run through all of the following definitions, models, as well as computational and statistical procedures for OTU table processing.

The Hill numbers (Chao et al. 2012, 2014a) for measuring alpha-diversity are defined as

$$ {}^{q}D = \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)}, \tag{1} $$

where D is the diversity in Hill numbers, q is the *order* number of diversity, S is the number of species (or OTUs), and p_i is the relative abundance of species (OTU) i. The diversity order (q) determines the sensitivity of the Hill numbers to the relative frequencies of species abundances. When $q = 0$, as species abundance is not involved in the calculation, ${}^{0}D$ is equal to the number of species, or *species richness* (S). When $q = 1$, the species abundance is involved in the calculation and ${}^{1}D$ represents the number of *typical* or *common* species in the community, which is equal to the exponential of Shannon entropy. Note that when $q = 1$, the ${}^{1}D$ is undefined and its limit form (Equation (2)) exists and is used to compute the Hill numbers. When $q = 2$, ${}^{2}D$ is more sensitive to the species with more abundant species (dominant species) than to common species, which is equal to the reciprocal of Simpson index. Generally, ${}^{q}D$ is the *diversity* of a community with $x = {}^{q}D$ equally abundant species. When $q = 3$ or higher, the computation of ${}^{3}D$ is weighted even more by more abundant (dominant) species and less by rarer species.

$$ {}^{1}D = \lim_{q \to 1} {}^{q}D = \exp\left(-\sum_{i=1}^{S} p_i \log(p_i) \right). \tag{2} $$

A primary reason why we adopt the Hill numbers in this study has to do with a consensus reached in a forum held by journal *Ecology*, i.e. the Hill numbers represent the most appropriate metric for measuring alpha-diversity (Ellison 2010, Chao et al. 2012, 2014a). A fundamental property of Hill numbers is that they are in the units of *numbers of species* or *species equivalents* (such as OTUs), which is made possible by the 'stratification' of the Hill numbers in terms of the diversity order (q), corresponding to different lev-

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

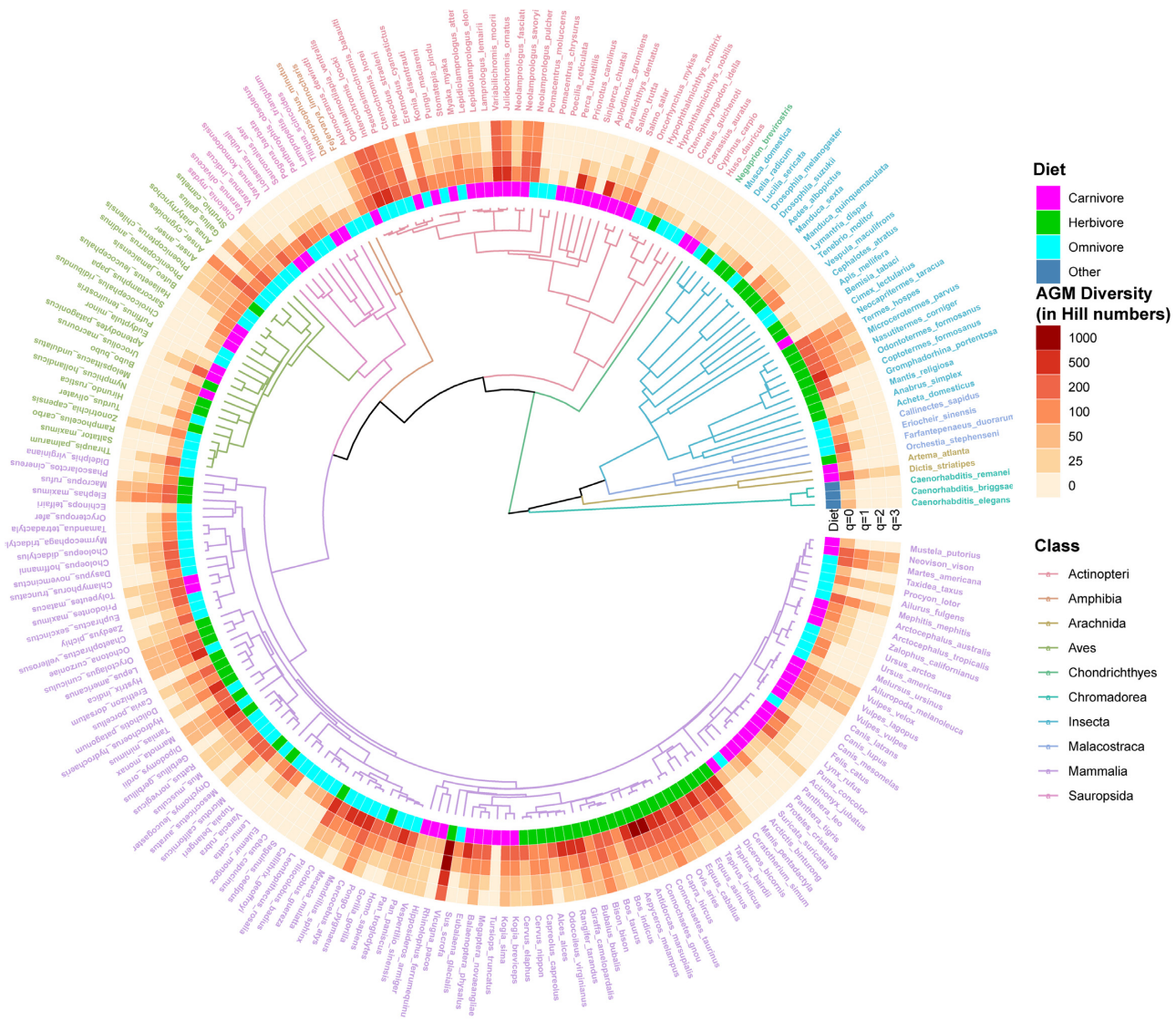

Figure 1. The phylogenetic tree (P-Tree) of 224 animal species, annotated with their gastrointestinal tract microbiome diversities: (i) Branches and species labels constitute a standard phylogenetic tree and were colored differently for each of the 10 animal classes (species labels were colored in terms of their class identities); (ii) Among the five colored bands, the inner-most band represents the diet type corresponding to each species, the other four bands constitute the heatmap of gut microbiome diversities (in the Hill numbers at diversity order $q = 0{\sim}4$) of the animal species.

els of weights determined by the SAD. As explained previously, while ${}^0D$ defaults to *species richness* and species abundance does not count at all, ${}^3D$ is strongly weighted by (in favor for) more dominant species. In other words, the Hill numbers are of the same 'standardized' unit (species or species equivalents) at a specific diversity order (q). This fundamental property with Hill numbers overcomes the inconsistencies in measuring/comparing diversities from choosing different diversity indexes such as species richness, Shannon entropy, or Simpson index, often encountered in practice. In the case of microbial diversity research, Hill numbers can be even more advantageous (Chiu and Chao 2016, Weiss et al. 2017).

However, the diversity measures (including the Hill numbers), estimated directly from its definitions can be biased. Even for the simplest diversity index, species richness, the estimation is highly sensitive to sample sizes, which is known as the 'sampling problem' in the literature and is deeply rooted in the nature of SAD. The SAD is usually highly skewed with long tail (representing for rare species) and different sampling efforts (sample sizes) can reveal different levels of long tail. To deal with the sampling problem, the rarefaction curve approach and its various derivations have been proposed to estimate species richness and also to display the possible asymptotic limit line. Chao et al. (2014b) formally developed the rarefaction approach to estimating Hill numbers, and we adopt their approach and iNEXT R-package (Hsieh et al. 2016) to compute the asymptotic estimates of Hill numbers. We standardized all computations of the asymptotic estimates to 10 000 of 16S rRNA reads to facilitate the comparisons of diversities among different samples from possibly different studies (which may have different sample sizes). Due to program limit, the asymptotic estimation was limited to diversity $q = 0$, 1, and 2. In prescreening experiments, we also computed the asymptotic Hill numbers based on the rarefaction to 20 000 reads and on the average reads per sample (average = 9800), respectively. The results from three reads schemes were rather similar, and we report the results from 10 000-reads rarefaction only since the 10 000-reads were

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

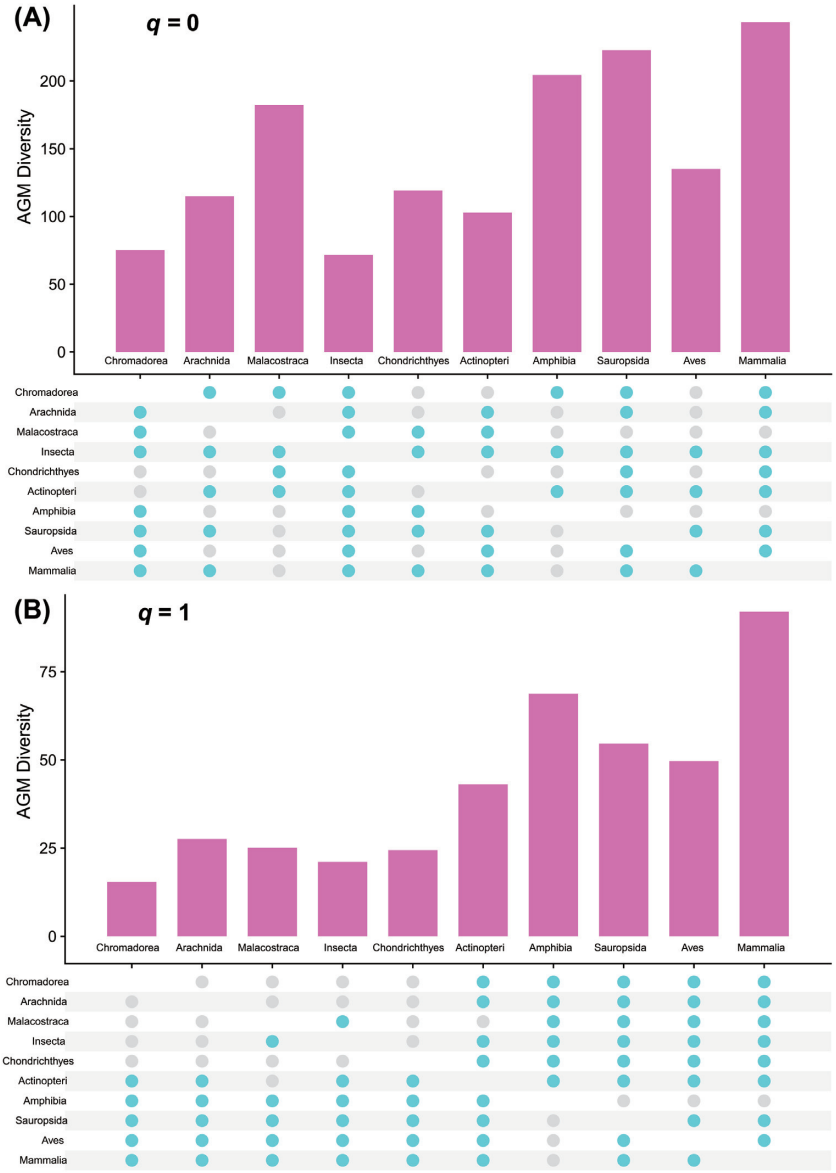

Figure 2. The average AGM diversities (in the Hill numbers at diversity order $q = 0$–3) of the 10 animal classes (top section), and the pair-wise class comparisons of the AGM diversities (bottom section): (i) the 10 animal classes are aligned on the x-axis in their phylogeny order; (ii) the circle in cyan color represents the class-pair with significant differences in their AGM diversities (P-value $\leq .05$; see Table S4 (Supporting Information) for the detailed numerical results). Fig. 2(A) for $q = 0$; Fig. 2(B) for $q = 1$; Fig. 2(C) for $q = 2$; and Fig. 2(D) for $q = 3$. When $q = 0$, species abundance does not weigh in the calculation of diversity, and obtained diversity is equal to species richness; when $q > 0$, species abundance weighs in.

also rather close to the average reads (9800) per sample in our study.

Table S2 (Supporting Information) lists the Hill numbers (i.e. the diversity at order $q = 0$–3) for each of the 4903 microbiome-samples collected from 318 animal species, and includes both direct estimates and rarefaction estimates (standardized to 10 000 reads as explained previously).

Modeling the relationship between Hill numbers, phylogeny, and diet types

We first try to establish a quantitatively significant relationship between the AGM-diversity (measured in the Hill numbers) and phylogeny (measured in the PT) of the animal host species across all four diet types (without distinguishing between diet types). We then try to establish the diversity–PT (DPT) relationship for each diet type to cross-verify possible relationships between diversity and phylogeny. In pre-experiments, we found that the following log-linear relationship fitted to the datasets well.

$$\ln(\text{Diversity}) = \ln(a) + b\ln(\text{PhylogeneticTimeline}). \quad (3)$$

The above log-linear model can be back-transformed into the following power law relationship:

$$\text{Diversity} = a(\text{PhylogenyTimeline})^b. \quad (4)$$

Results

We organize the results as (i) the effects of phylogeny and/or diet types on the AGM-diversity, both alone and jointly, through statistical tests; (ii) quantitative modeling of the relationship between PT of animal species and AGM-diversity. To prepare for the modeling, we first draw Fig. 1 to show the phylogenetic tree (P-Tree)

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

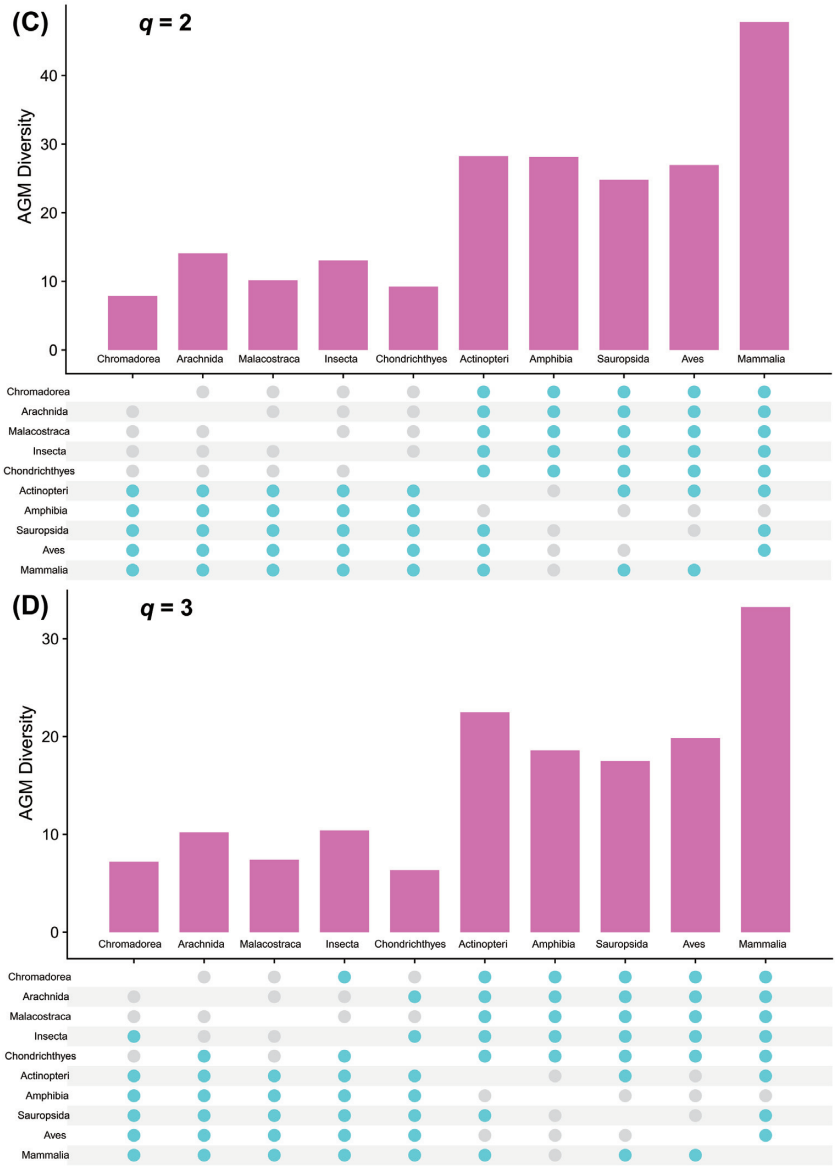

Figure 2. *Continued.*

of 224 animal species, annotated with their AGM diversities (see Table S6 (Supporting Information) for the diversities). The PT information was available for the 224 species in the P-Tree only (see Fig. 1), and Table S1 (Supporting Information) listed a summary on those 224 species. Although separately reported for the results from statistical tests and DPT modeling, both the results are interdependent. In fact, the latter modeling approach further deepens the findings from the statistical tests.

The effects of phylogeny and diet types on AGM-diversity detected by statistical tests

Based on the statistical tests designed for detecting the phylogeny and diet effects, we found that both host-phylogeny and diet can influence the diversity of AGM according to Sheirer–Ray–Hare test (i.e. the two-way nonparametric ANOVA, Table S3, Supporting Information) and Kruskal–Wallis one-way nonparametric ANOVA (P-value $\leq$ .001, Fig. 2 and Table S4 (Supporting Information) for phylogeny, Fig. 3 and Table S5 (Supporting Information) for diet type). Table 2 further summarizes Tables S3–S5 (Supporting Information) and listed the percentage of the compared pairs with significant difference in their AGM-diversity in terms of the class effects (top section) and diet effects (bottom section), respectively. What slightly complicates the exposition of the results is that the effects of phylogeny and/or diet types on AGM-diversity may vary with the diversity orders ($q = 0$–3). As explained in the section of 'Material and methods,' we measure the diversity in Hill numbers across a series of diversity orders (q): specifically, the number of species in the community or species richness ($q = 0$) when species abundance does not weigh in computing the Hill number at all; the number of typical or common species ($q = 1$) when species abundance weigh proportionally in computing the Hill number; the number of dominant species ($q = 2$) when more abundant species (dominant species) weigh in more than rarer species in computing the Hill number. That is, the higher the diversity order (q) is, the heavier weights from high abundant species in computing the corresponding Hill number.

Specifically, at diversity order $q = 0$ (i.e. species richness), the class of *Insecta* has significantly lower diversity (species richness)

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

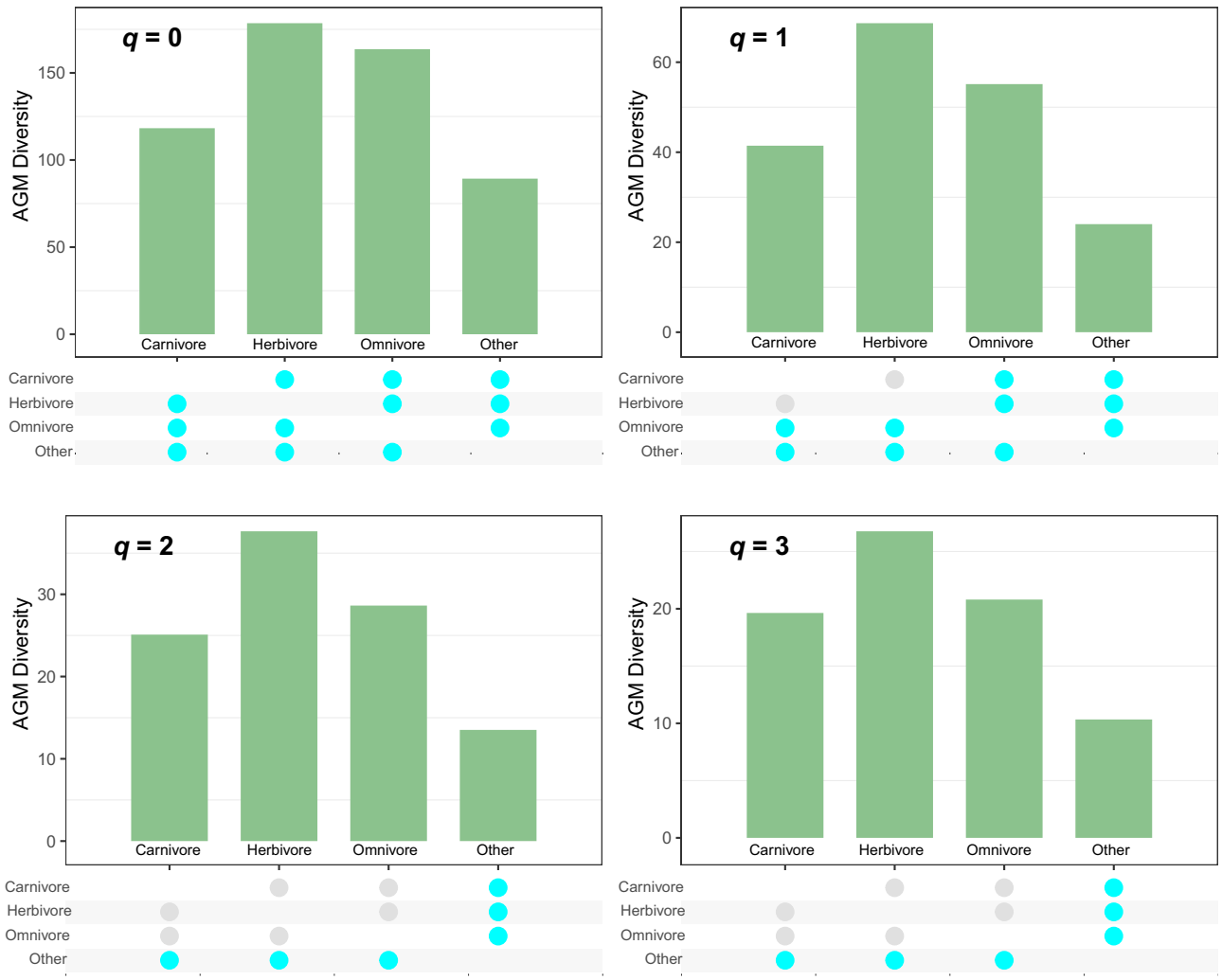

Figure 3. The average AGM diversities (in the Hill numbers at diversity order $q = 0$–3) of the four diet types (top section), and the pair-wise diet comparisons of the AGM diversities (bottom section): the circle in cyan color represents the diet-pair with significant differences in their AGM diversities (P-value $\leq .05$; see Table S5 (Supporting Information) for the detailed numerical results). Figure 3(A) for $q = 0$; Fig. 3(B) for $q = 1$; Fig. 3(C) for $q = 2$; and Fig. 3(D) for $q = 3$. When $q = 0$, species abundance does not weigh in the calculation of diversity, and obtained diversity is equal to species richness; when $q > 0$, species abundance weighs in.

than all other nine classes. At other diversity orders ($q = 1$–2), their diversities are still lower than vertebrates (except *Chondrichthyes*) but can be higher than some invertebrates (Tables S2 and S4, Supporting Information). Overall, more than $\frac{1}{2}$ of the animal class pairs (53%–89%) exhibited significant differences in their microbial diversity. In terms of the diet type, at diversity order $q = 0$ (species richness) all four diet types exhibited significant differences in the AGM-diversity. Significant differences in the AGM-diversity were also detected in 83% of diet pairs when $q = 1$, and in 50% of diet pairs when $q = 2$ and 3. In addition, there were significant differences between the 'other' type (i.e. the diet type of unclear diet) and carnivorous, herbivorous, and omnivorous diet types at all four diversity orders ($q = 1$–3). Given the undefined nature of the 'other' type, we do not highlight the 100% difference associated with the 'other' type.

Modeling the relationships between AGM-diversity and PT

We first built the AGM-DPT power law models (Equations (3) and (4)) by pooling together the four diet types, which allow us to focus on the general trend of DPT patterns. Specifically, Fig. 2 displays a not fully consistent trend: the diversity (H) seems to increase from ancient to more modern species (i.e. from larger PT to smaller PT), especially for $q = 1$~3. For example, vertebrates, except for fish, generally possess higher diversity than invertebrates. Not surprisingly, Mammalia exhibited the highest AGM-diversity out of all of the animal classes. To further quantify this trend rigorously, the log-linear model (Equation (3)) was fitted to test the relationship between the AGM-diversity and host PT of animal species, i.e. the previously defined DPT relationship (Equations (3) and (4)). Table 3 below showed the detailed model-fitting results. Table 3(A) (the top section of Table 3) lists the four models for each diversity order ($q = 0$–3) by using the AGM data of all animal species (without distinguishing between diet types). Table 3(B) (the bottom section of Table 3) lists the 12 models for the 12 combinations of the diversity order ($q = 0$–3) and diet type by distinguishing between diet types (3 diet types × 4 diversity orders = 12 combinations). As shown by the P-value ($< .000$) and the all-negative slope (b) of the log-linear models in Table 3, the previously observed trend indeed holds in general. That is, the DPT relationship between the AGM-diversity and host PT is negatively log-linear, or negatively linear on the log-scale. Equivalently, the log–linear relationship is also a power-law

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

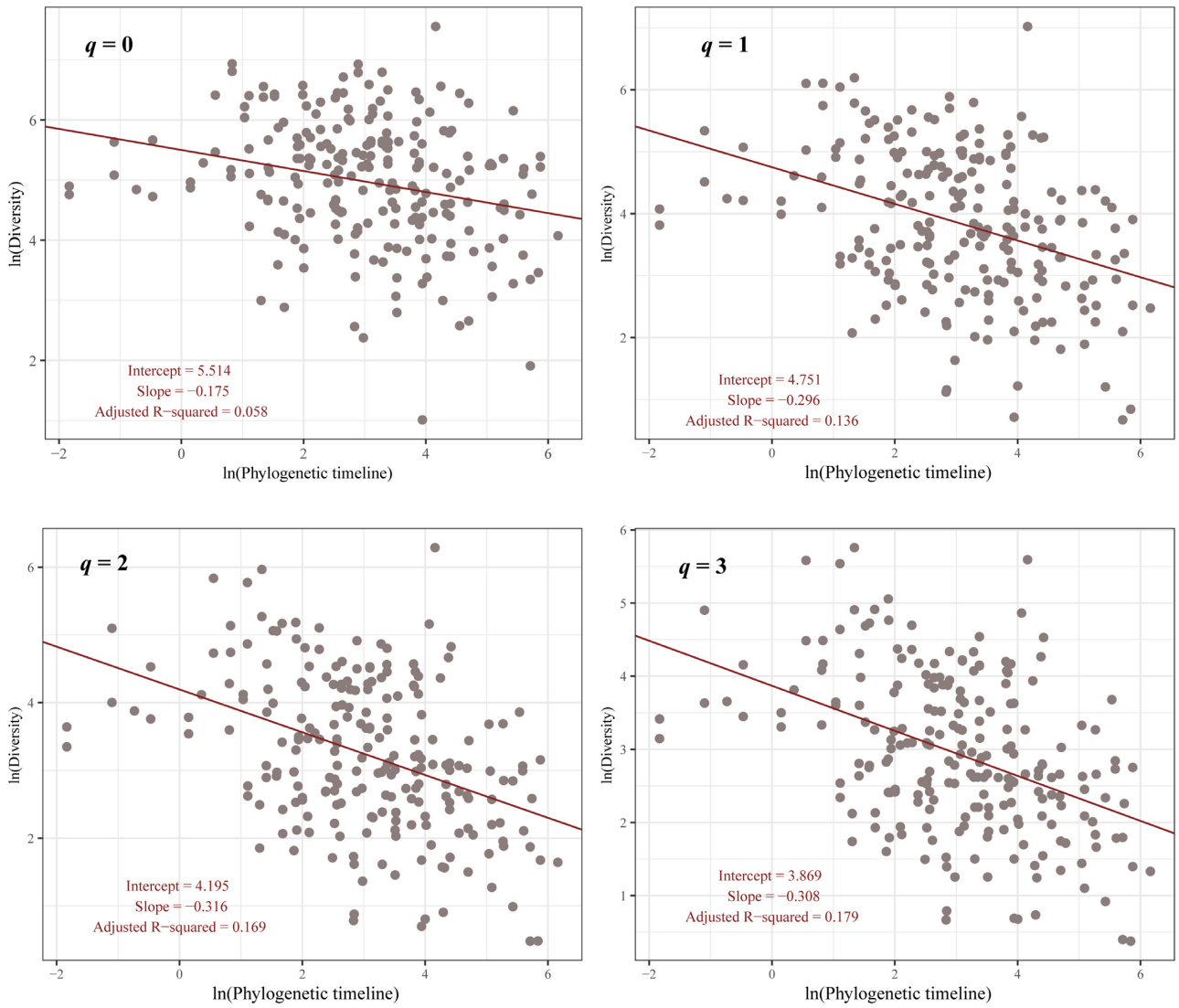

Figure 4. The log-linear DPT model (or linear model on the log-scale, equivalent with power-law model) between the AGM-diversity and the PT of animal species (the *P*-value from *F*-test for all four models at diversity order $q = 0$–3 satisfied with $P < .01$), i.e. the unified DPT models across diet types.

relationship, because Equation (3) is simply a log-transformation of the power-law model (Equation (4)). Figure 4 plots the graphs of DPT log–linear relationships for each diversity order $q = 0$–3 with all diet types pooled together based on Table 3(A).

We further built diet type-specific DPT models, which allow us to obtain more detailed insights on the diet effects on the DPT patterns. Figure 5 plots the DPT relationships for each diet type separately based on Table 3(B). Figures 4 and 5 show the same trend previously identified, i.e. the inverse log–linear relationship between diversity and PT regardless of diet types. As a side note, to ensure the robustness of DPT modeling, we only included the AGM datasets of 224 animal species (Table S1, Supporting Information), for which we could get reliable PT data. Furthermore, we computed nonparametric Spearman's correlation coefficients (listed in the left side of Table 3) to cross-verify the log-linear modeling of the DPT relationship, and the nonparametric Spearman's correlation coefficients indeed cross-verified the negative correlation between the diversity and PT.

The bottom section (Table 3B) of Table 3 lists *diet-specific* DPT power law model. Here, we compare three different diet types (the 'other' type was excluded due to significantly small sample sizes with the other three types) in terms of their scaling parameter (b). Note that the order of scaling parameter (b) in Table 3 here, can be different from the order of AGM-diversity displayed in Fig. 3 (also in Table S5, Supporting Information), because the former describes the change (scaling) of diversity across species, but the latter describes the actual magnitude of diversity *per se*, similar to the *acceleration rate* of velocity vs. velocity in physics. While Fig. 3 (Table S5, Supporting Information) shows the order of diversity is *herbivore>omnivore>carnivore* (the counterpart of velocity in the velocity–acceleration rate analogy), the relationships (order) of diversity scaling (b) (the counterpart of acceleration rate in the analogy) are much more complex. Again the patterns seem different when $q = 0$ and $q > 0$, as in previous all-species DPT relationship. When $q = 0$ or for species richness, herbivores exhibited the highest diversity scaling parameter (b), and nearly $\frac{1}{2}$ larger than carnivore or omnivore. That is, herbivores have the most dramatic changes across animal species in terms of species richness ($q = 0$). At diversity order $q = 1$, i.e. or in terms of *common* species, herbivores still possess slightly higher scaling (change) rates across animal (herbivore) species than omnivores and carnivores. At higher diversity order $q = 2$ and 3, or in terms of dominant species (more

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

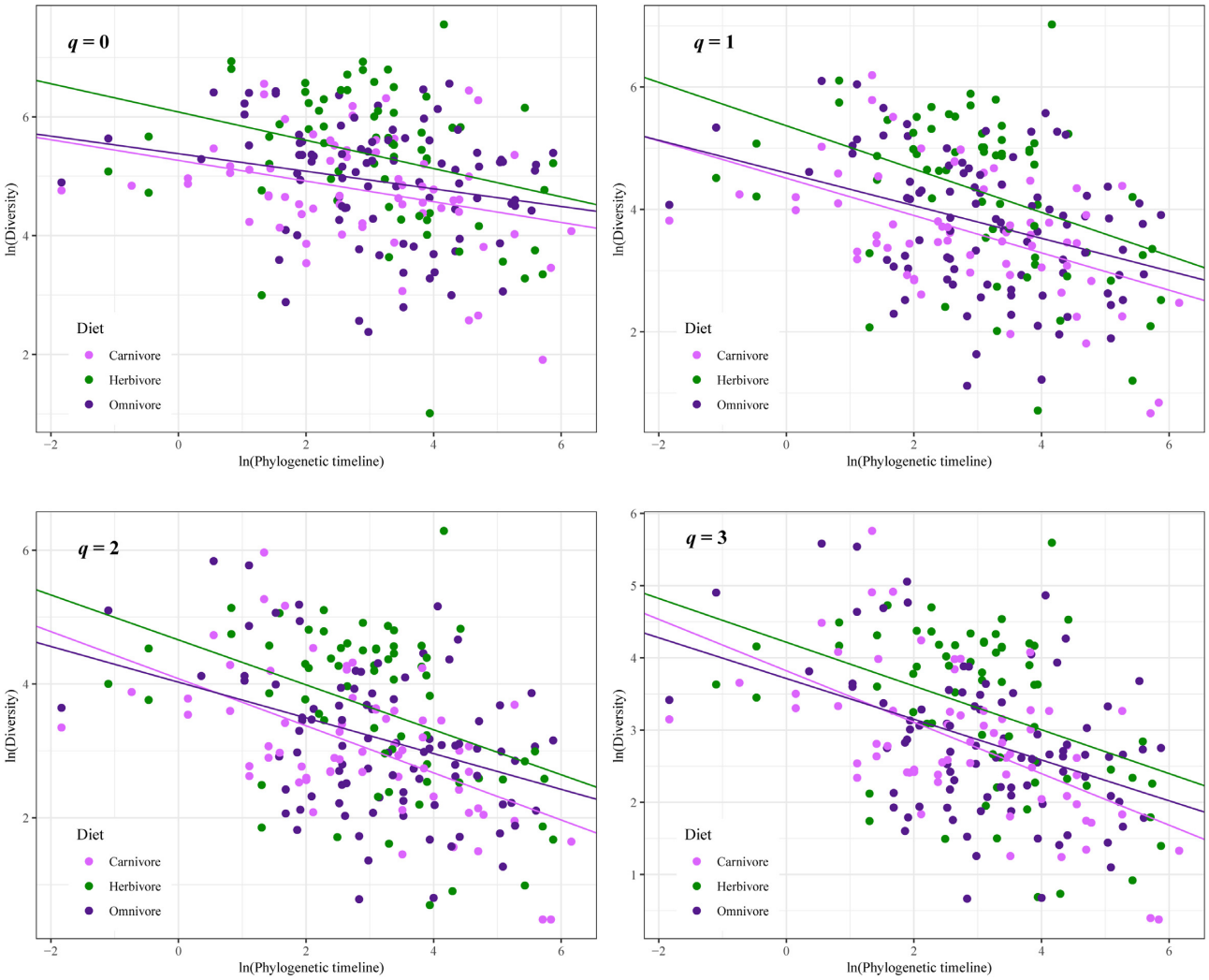

Figure 5. The log-linear DPT model (or linear model on the log-scale, equivalent with power-law model) between the AGM-diversity and PT of the animal species from each of the three diet types (carnivore, herbivore, and omnivore), respectively (*P*-values from *F*-test for all 12 models: $P < .05$), i.e. the diet-specific DPT models.

Table 2. The percentage (%) of the pair-wise class (or diet) comparisons with significant differences in the AGM-diversity (in Hill numbers) for each of the 10 classes (or each of the four diet types), summarized from Tables S4 and S5 (Supporting Information), respectively.

Class	$q = 0$	$q = 1$	$q = 2$	$q = 3$	Average (%)
		Permutation tests for the effects of classes			
Chromadorea	77.8% (7/9)	55.6% (5/9)	55.6% (5/9)	66.7% (6/9)	**63.9**
Arachnida	55.6% (5/9)	55.6% (5/9)	55.6% (5/9)	66.7% (6/9)	**58.4**
Malacostraca	44.4% (4/9)	55.6% (5/9)	55.6% (5/9)	55.6% (5/9)	**52.8**
Insecta	100% (9/9)	66.7% (6/9)	55.6% (5/9)	77.8% (7/9)	**75.0**
Chondrichthyes	55.6% (5/9)	55.6% (5/9)	55.6% (5/9)	77.8% (7/9)	**61.2**
Actinopteri	66.7% (6/9)	88.9% (8/9)	88.9% (8/9)	77.8% (7/9)	**80.6**
Amphibia	33.3% (3/9)	66.7% (6/9)	55.6% (5/9)	55.6% (5/9)	**52.8**
Sauropsida	77.8% (7/9)	88.9% (8/9)	77.8% (7/9)	77.8% (7/9)	**80.6**
Aves	44.4% (4/9)	88.9% (8/9)	77.8% (7/9)	66.7% (6/9)	**69.5**
Mammalia	77.8% (7/9)	88.9% (8/9)	88.9% (8/9)	88.9% (8/9)	**86.1**
Average	**63.34%**	**71.14%**	**66.70%**	**71.14%**	
		Permutation tests for the effects of diet types			
Diet	**$q = 0$**	**$q = 1$**	**$q = 2$**	**$q = 3$**	**Average** (%)
Carnivore	100% (3/3)	66.7% (2/3)	33.3% (1/3)	33.3% (1/3)	**58.33**
Herbivore	100% (3/3)	66.7% (2/3)	33.3% (1/3)	33.3% (1/3)	**58.33**
Omnivore	100% (3/3)	100% (3/3)	33.3% (1/3)	33.3% (1/3)	**66.65**
Other	100% (3/3)	100% (3/3)	100% (3/3)	100% (3/3)	**100**
Average	**100%**	**83.35%**	**49.98%**	**49.98%**	

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

Table 3. The relationships between the AGM diversity and PT of the animal species, i.e. the DPT relationships.

A. The DPT models for all species

		Spearman's correlation		Log-linear model (power-law model)			
Diversity order	Taxon	*R*	*P*-value	ln(*a*)	*b*	Adjusted R^2	*P*-value
$q = 0$	All species	−0.266	.000	5.514	−0.175	0.058	.000
$q = 1$	All species	−0.374	.000	4.751	−0.296	0.136	.000
$q = 2$	All species	−0.401	.000	4.195	−0.316	0.169	.000
$q = 3$	All species	−0.413	.000	3.869	−0.308	0.179	.000

B. The DPT models for each of the four diet types

Diversity order	Diet type	Spearman's correlation		Log-linear model (power law model)			
		R	*P*-value	ln(*a*)	*b*	Adjusted R^2	*P*-value
$q = 0$	Carnivore	−0.290	.019	5.266	−0.174	0.077	.014
	Herbivore	−0.366	.003	6.081	−0.238	0.073	.017
	Omnivore	−0.265	.011	5.379	−0.148	0.038	.036
$q = 1$	Carnivore	−0.419	.001	4.511	−0.305	0.215	.000
	Herbivore	−0.430	.000	5.368	−0.354	0.148	.001
	Omnivore	−0.349	.001	4.596	−0.267	0.113	.001
$q = 2$	Carnivore	−0.507	.000	4.079	−0.352	0.293	.000
	Herbivore	−0.398	.001	4.659	−0.336	0.151	.001
	Omnivore	−0.336	.001	4.027	−0.286	0.14	.000
$q = 3$	Carnivore	−0.547	.000	3.821	−0.356	0.314	.000
	Herbivore	−0.379	.002	4.216	−0.303	0.144	.001
	Omnivore	−0.341	.001	3.714	−0.282	0.149	.000

abundant than common species), the previous trend reversed and carnivores exhibited slightly higher diversity scaling rates across species than herbivores and omnivores.

Conclusions and discussion

This meta-analysis identified associations between AGM-diversity and phylogeny (evolutionary history) and diet types of host animals. The relationship between AGM-diversity (measured in Hill numbers) and PT, i.e. DPT, is negatively log-linear, i.e. negatively and linearly correlated on the log-scale. Equivalently, the DPT relationship follows a power law with negative scaling exponent since the latter (Equation (4)) is the back-transformation of the former (Equation (3)). The power-law relationship indicates that the DPT relationship is asymmetrical or highly skewed with a long tail, suggesting that a small handful of taxa carry disproportionally large weights to the evolution of the diversity pattern than the majority of taxa. The power-law relationship also explains why *Mammalia* exhibited high AGM-diversity than the other classes investigated in this study, which cover all six classes of the vertebrates and the four major classes of the invertebrates. Similarly, it explains why *Insecta* exhibit significantly lower AGM species richness (diversity order $q = 0$) than all other nine classes. For diversity order $q = 1$ and 2, the AGM-diversity of *Insecta* is still lower than those of vertebrates (except for *Chondrichthyes*) but can be higher than the diversity of some invertebrates. These two observations highlight highly uniqueness of insects and mammalia in the AGM DPT relationships. We postulate that the microbiomes could have played critical roles for their co-evolutions with their respective hosts, which, in turn, explains the uniqueness of AGM DPT relationships of both the classes. However, as mentioned previously in the instruction section, it may be the case that only some 'driver' microbes are involved in co-evolution, and those 'passenger' microbes may simply 'piggyback' on their hosts.

Existing studies such as Ley et al. (2008), Muegge et al. (2011), Delsuc et al. (2014), Vital et al. (2015), and Amato et al. (2019) have revealed the diet effects on the diversity of animal gut microbiomes to different levels. Our study demonstrates prevalent diet effects on the AGM-diversity, 100% on species richness (i.e. Hill numbers at diversity order $q = 0$), 83% on the diversity of typical or common species (i.e. Hill numbers at diversity order $q = 1$), and 50% on the diversity of dominant species (i.e. Hill numbers at diversity order $q = 2$ and 3).

While some existing studies suggested that host-phylogeny has rather limited effects on the gut microbiome (e.g. Degnan et al. 2012, Carmody et al. 2015, Martinson et al. 2017, Gomez et al. 2019, Trevelline et al. 2020), other studies confirmed significant roles of evolutionary history (e.g. Moeller et al. 2014, 2016, Gaulke et al. 2018, Li et al. 2018, Sherrill-Mix et al. 2018, Amato et al. 2019, Youngblut et al 2019, Song et al. 2020, Rojas et al. 2021). Our study not only confirmed the effects of phylogeny revealed by previous studies, but also established quantitative log-linear (power-law) models between diversity and PT, indicating the *predictive* nature of the relationships. Still some other existing studies such as Groussin et al. (2017) suggested that host diet and evolutionary history both significantly influenced the structure of gut microbiome, but aspects of effects could be different. Our diet-specific DPT models also revealed similar intricacies, not necessarily from different lineage perspective, but from different levels of diversity orders, which stratify the weights of species abundances on diversity (Hill numbers). The diversity order can be considered as proxies for different *segments* of the SAD. Different segments of SAD represent for species with different level of species abundances, e.g. $q = 2$ corresponding to dominant species, and $q = 1$ corresponding to common species. Such segmentation or complexity is a reality in the DPT–diet relationships, and such complexity is likely due to the interdependency between *phylogeny* and *diet*. As we argued previously, diet adaptation *per se* is product of evolution, and therefore, it might be unrealistic to disentangle the interdependency between phylogeny and diet types. Despite this likely insurmountable challenge, the DPT models we obtained here, to the minimum, offer a tool to recognize and dissect the complexity.

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

In addition, the *all-animal species* DPT power-law model (top section of Table 3) suggests that species richness ($q = 0$) has rather different scaling parameter (b) (or the slope of log-linear model), and the magnitude is almost doubled in the higher diversity orders (q = 1–3). This indicates that the change of total microbial species numbers (richness) across animal species is much less dramatic than the change of diversity at high diversity orders ($q > 0$), when species abundances are taken into accounts. While this finding should not be surprised, since it simply says that variability in microbial species abundance is higher than the variability in the number of species, which should be the case in general (first principle), the finding highlights the importance of choice in diversity measures. The Hill numbers we adopted do not face choice issue because the Hill numbers present a comprehensive diversity profile stratified at different diversity orders (q = 0–3), and any comparisons are made on the same diversity order.

One limitation of our study is the failure to consider the effects of other potentially important factors such as host geography (Moeller et al. 2017) and habitats on the gut microbiome diversity. Future studies are needed to further address the effects of these more complex environmental factors on animal gut microbiome diversities. The approaches demonstrated in this study can be equally useful for analyzing their effects. Finally, a caution is worthy of mention regarding the uncertainty in estimating the PT of various animal species, which may impact the goodness-of-fitting of our DPT models. In extreme cases, the PT information for some animal species is currently unavailable. In addition, the uncertainty (reliability) of the PT for different species may vary, but the reliability should improve steadily in future with the rapid progress in genomic and metagenomic studies. For this reason, we caution that the current DPT models built in this article should be treated as qualitative, exploratory tool, rather than quantitative and precise, predictive models.

Finally, a somewhat unique feature of this study is the adoption of Hill numbers as microbiome diversity metrics, which has an advantage of stratifying diversities according to different orders (q), i.e. weighted differently by the rarity (commonness) of microbial species or OTUs. For example, at higher diversity order (q = 2 and 3 in this study), the diversity represented by Hill numbers is weighted more by more abundant species. As mentioned previously, we postulate that microbes of AGM are not 'born' equal: some microbes (and the genes they carry) are 'passengers', rather than 'drivers', and 'passengers' are not involved in co-adaptation or co-evolution processes with their hosts as 'drivers' do. With this hypothesis, the concept of holobiont (hologenome) may be more appropriate if those 'passenger' microbes are excluded in consideration of their lack of participation in co-adaptation or co-evolution. Although we realize that there is not a simple measure to distinguish driver species from passenger species yet, rare species might be more likely, than common or dominant species, to be passenger species. If this conjecture is true, the diversity or Hill numbers at higher diversity-orders may better reflect the contribution of driver or resident microbes. However, we cannot make a similar claim for passenger microbes because the diversity at lower diversity orders is not weighted by rare species alone. Significant challenges exist for characterizing the proposed 'drivers' vs. 'passenger' dichotomy (hypothesis), but the payoff, in our opinion, is huge.

Data availability

Table 1 and Table S1 (Supporting Information; an Excel table) in the OSI-Excel (Online Supplementary Information Excel file) presented detailed sample information including the data accession numbers. The data were already available in public domain and no ethic or administrative approval is applicable.

Supplementary data

Supplementary data are available at *FEMSEC* online.

Authors' contribution

Z.M. and P.S. designed the study, W.L. and Z.M. performed the data analysis, All authors participated in the interpretation of the results and revisions of the manuscript; Z.M. wrote the paper. All authors approved the submission.

Funding

This study received funding from the following sources: a National Natural Science Foundation (NSFC) grant (number 31970116) on 'Medical Ecology of Human Microbiome;' The Cloud-Ridge Industry Technology Leader Award. The funders played no roles in interpreting the results.

Acknowledgments

We deeply appreciate the two anonymous reviewer experts and editor Dr. Cindy Nakatsu for their exceptionally meticulous and patient comments, suggestions, and insights, which helped to improve our work significantly.

Conflicts of interest. None declared.

References

Amato KR, Sanders G, J. S *et al.* Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes. *ISME J* 2019;**13**:576–87. DOI: 10.1038/s41396-018-0175-0.

Bolyen E, Rideout JR, Dillon MR *et al.* QIIME 2: reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints* 2018;**6**:e27295v2.

Brucker RM, Bordenstein SR. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *nasoni a*. *Science* 2013;**341**:667–9.

Carmody RN, Gerber GK, Luevano JM *et al.* Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* 2015;**17**:72–84.

Chao A, Chiu CH, Hsieh TC. Proposing a resolution to debates on diversity partitioning. *Ecology* 2012;**93**:2037–51.

Chao A, Chiu CH, Jost L. Unifying species diversity, phylogenetic diversity, functional diversity and related similarity and differentiation measures through hill numbers. *Ann Rev Ecol Evol Syst* 2014a;**45**:297–324.

Chao A, Gotelli NG, Hsieh TC *et al.* Rarefaction and extrapolation with hill numbers: a frame- work for sampling and estimation in species biodiversity studies. *Ecol Monogr* 2014b;**84**:45–67. https://github.com/JohnsonHsieh/iNEXT.

Chiu CH, Chao A. Estimating and comparing microbial diversity in the presence of sequencing errors. *PeerJ* 2016;**4**:e1634. DOI: 10.7717/peerj.1634.

Degnan PH, Pusey AE, Lonsdorf EV *et al.* Factors associated with the diversification of the gut microbial communities within chimpanzees from gombe national park. *Proc Natl Acad Sci* 2012;**109**:13034–9.

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024

Delsuc F, Metcale JL, Parferey LW *et al.* Convergence of gut microbiomes in myrmecophagous mammals. *Mol Ecol* 2014;**23**:1301–17.
Ellison AM. Partitioning diversity. *Ecology* 2010;**91**:1962–3.
Gaulke CA, Arnold HK, Humphreys IR *et al.* Ecophylogenetics clarifies the evolutionary association between mammals and their gut microbiota. *Mbio* 2018;**9**:e01348–18.
Gomez A, Sharma AK, Mallott EK *et al.* Plasticity in the human gut microbiome defies evolutionary constraints. *Msphere* 2019;**4**:e00271–19. DOI: 10.1128/mSphere .00271-19.
Groussin M, Mazel F, Sanders JG *et al.* Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat Commun* 2017;**8**:14319.
Hsieh TC, Ma KH, Chao A. iNEXT: an r package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol Evol* 2016;**7**:1451–6.
Kumar S, Stecher G, Suleski M *et al.* TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol* 2017;**34**:1812–9.
Ley RE, Hamady M, Lozupone C *et al.* Evolution of mammals and their gut microbes. *Science* 2008;**320**:1647–51.
Li J, Zhan S, Liu S *et al.* Divergence of fecal microbiota and their associations with host phylogeny in cervinae. *Front Microbiol* 2018;**9**:1–11.
Ma ZS, Li LW, Gotelli NJ. Diversity-disease relationships and shared species analyses for human microbiome-associated diseases. *ISME J* 2019;**13**:1911–9. https://www.nature.com/articles/s41396-019-0395-y.
Ma ZS, Li LW. Measuring metagenome diversity and similarity with hill numbers. *Mol Ecol Resour* 2018;**18**:1339–55.
Ma ZS. Measuring microbiome diversity and similarity with Hill numbers. Chapter 8, *In: Metagenomics*, Elsevier, 2017.
Martinson VG, Douglas AE, Jaenike J. Community structure of the gut microbiota in sympatric species of wild Drosophila. *Ecol Lett* 2017;**20**:629–39.
Matschiner M, Böhne A, Ronco F *et al.* The genomic timeline of cichlid fish diversification across continents. *Nat Commun* 2020;**11**:5895.
Mazel F, Davis KM, Loudon A *et al.* Is host filtering the main driver of phylosymbiosis across the tree of life?. *MSystems* 2018;**3**:1–15.
Moeller AH, Caro-Quintero A, Mjungu D *et al.* Cospeciation of gut microbiota with hominids. *Science* 2016;**353**:380–2.
Moeller AH, Li Y, Ngole EM *et al.* Rapid changes in the gut microbiome during human evolution. *Proc Natl Acad Sci* 2014;**111**:16431–5.
Moeller AH, Suzuki TA, Lin D *et al.* Dispersal limitation promotes the diversification of the mammalian gut microbiota. *Proc Natl Acad Sci* 2017;**114**:52.
Muegge DB, Kuczynski J, Knights D *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 2011;**332**:970–4.
Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 2004;**20**:289–90.
Roja CA, Rami´rez-Barahona S, Holekamp KE *et al.* Host phylogeny and host ecology structure the mammalian gut microbiota at different taxonomic scales. *Anim Microbiome* 2021;**3**:33.
Rosenberg E, Sharon G, Zilber-Rosenberg I. The hologenome theory of evolution contains lamarckian aspects within a darwinian framework. *Environ Microbiol* 2009;**11**:2959–62.
Rosenberg E, Zilber-Rosenberg I. The hologenome concept of evolution after 10 years. *Microbiome* 2018;**6**:78.
Sherrill-Mix S, McCormick K, Lauder A *et al.* Allometry and ecology of the bilaterian gut microbiome. *Mbio* 2018;**9**:e00319–18.
Song SJ, Sanders JG, Delsuc F *et al.* Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. *MBio* 2020;**11**:1–14.
Theis KR, Dheilly NM, Klassen JL *et al.* Getting the hologenome concept right: an eco-evolutionary framework for hosts and their microbiomes. *Msystems* 2016;**1**:e00028–16.
Trevelline BK, Sosa J, Hartup BK *et al.* A bird's-eye view of phylosymbiosis: weak signatures of phylosymbiosis among all 15 species of cranes. *Proc R Soc B Biol Sci* 2020;**287**:20192988.
Vital M, Gao J, Rizzo M *et al.* Diet is a major factor governing the fecal butyrate-producing community structure across mammalia, aves and reptilia. *ISME J* 2015;**9**:832–43.
Weiss S, Xu ZZ, Peddada S *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 2017;**5**:27.
Wolfe JM, Breinholt JW, Crandall KA *et al.* A phylogenomic framework, evolutionary timeline and genomic resources for comparative studies of decapod crustaceans. *Proc Biol Sci* 2019;**286**:20190079.
Youngblut ND, Reischer GH, Walters W *et al.* Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nat Commun* 2019;**10**:2200.
Yu G, Smith DK, Zhu H *et al.* ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;**8**:28–36.

Downloaded from https://academic.oup.com/femsec/article/98/2/fiac021/6534254 by Kennedy School of Government Library user on 21 June 2024