Diversity scaling of human digestive tract (DT) microbiomes: the intra-DT and inter-individual patterns

Hongju Chen (daisyhjchen@outlook.com)
University of the Chinese Academy of Sciences https://orcid.org/0000-0001-6715-9217
Bin Yi
Honghe University
Xia Xu
Honghe University
Qiang Liu
Honghe University
Lin Dai
Kunming University of Science and Technology
ZS (Sam) Ma
University of the Chinese Academy of Sciences

Research article

Keywords: Inter-individual, intra-DT, heterogeneities, diversity-area relationship (DAR), Diversity area relationship (DAR) profile, Potential diversity, Dark diversity

Posted Date: July 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-46116/v1

Diversity scaling of human digestive tract (DT) microbiomes: the intra-DT and inter-individual patterns

Hongju (Daisy) Chen[123#] Bin Yi[3#] Xia Xu[3] Qiang Liu[3]
Lin Dai[4] ZS (Sam) Ma[125*]

[1]Computational Biology and Medical Ecology Lab,
State Key Laboratory of Genetic Resources and Evolution,
Kunming Institute of Zoology,
Chinese Academy of Sciences, Kunming, China

[2]Kunming College of Life Sciences,
University of Chinese Academy of Sciences, Kunming, China

[3]College of Mathematics,
Honghe University, Yunnan, China

[4]Faculty of Science
Kunming University of Science and Technology, Kunming, China

[5]Center for Excellence in Animal Evolution and Genetics
Chinese Academy of Sciences, Kunming, China

#These authors (HC & BY) made equal contributions.
*Contact email: ma@vandals.uidaho.edu

Abstract

The human gut microbiome has been extensively studied, but its diversity scaling (changes) along the DT (digestive tract) as well as their inter-individual heterogeneities have not been adequately addressed in our opinion. Here we fill the gap by applying the diversity-area relationship (DAR), a recent extension to the classic species-area relationship (SAR) in biogeography, by reanalyzing the dataset of over 2000 16s-rRNA microbiome samples obtained from 10 DT sites of over 200 individuals. We sketched out the biogeography "maps" for each of the 10 DT sites by cross-individual DAR analysis, and the intra-DT distribution pattern by cross-DT site DAR analysis. Regarding the inter-individual biogeography, it was found that all DT sites have the invariant scaling parameter—all sites possessing the same diversity change rate across individuals, but most sites have different *potential* diversities. In terms of the genus richness, an average individual hosts approximately 20% of the population-level genus richness (total bacterial genus of a human population). In contrast, in terms of community biodiversity, the percentages of individual *vs*. population may exceed 90%. This suggests that the differences between individuals in their DT microbiomes are predominantly in the composition of bacterial species, rather than how their abundances are distributed (*i.e*., biodiversity). Regarding the intra-DT patterns, the scaling parameter is larger—suggesting that the intra-DT biodiversity changes are more dramatic than inter-individual changes. On average, each site contains 21%-36% of genus diversity of the whole DT, and the percentages are even higher at the higher taxon levels.

Keywords: Inter-individual; intra-DT; heterogeneities; diversity-area relationship (DAR); Diversity area relationship (DAR) profile; Potential diversity; Dark diversity

Introduction

There have been extensively and intensive studies on various aspects of human gut microbiomes in the last decade and so particularly on their diversities (for example, HMP Consortium 2012a, 2012b, Lozupone et al. 2012, Segata *et al.* 2012). However, certain information on the diversity scaling (changes) across individuals or across various sites of an individual's DT (digestive tract) seems still missing. For example, to what extent, an individual's gut microbial species richness (number of species) can represent for a human population? What about species diversity (such as Shannon evenness index) if individual *vs*. population is compared? How is the microbiome distributed along one's DT, in terms of species richness or diversity? How about the scaling relationships on higher taxon level, such as phylum, class, order, and family? Similar questions are traditionally investigated in biogeography. One of the most well-known ecological laws in

the field is the so-termed species-area relationship (SAR), which was first discovered in the 19th century (Watson 1835) and has been extensively investigated since 1960s (Preston 1960, Connor & McCoy 1979, Rosenzweig 1995, Lomolino 2000, Tjørve 2008, 2009, Drakare et al. 2006, Harte et al. 2009, He & Hubbell 2011, Sizling et al. 2011, Storch et al. 2012, Triantis et al. 2012, Helmus et al. 2014). It is considered as "ecology's most general, yet protean pattern" by Lomolino (2000) and Whittaker & Triantis (2012). In practice, SAR has become one of the most important theory and mode in conservation biology and biodiversity protection. This is because SAR can be applied to establish a simple function relationship between the number of species (species richness) and the area of a region. It was found that SAR typically follow a simple power function in the form of $S=cA^z$, where S is the number of species, A is size of area, z and c are SAR parameters. In particular, z is termed SAR scaling parameter, and it is a measure of change (increase) rate of species over area.

It is generally recognized that, the number of species or formally species richness, is a rather convenient but very rough measure of biodiversity. This is because biodiversity is obviously strongly influenced by both species richness and species abundances of individual species. Several biodiversity metrics that consider both species numbers and abundances have been proposed and widely used in ecology since 1960s. A recent consensus has been that the Hill numbers, which were first introduced by Hill (1973) into ecology but did not receive significant attention until recently, offer the most appropriate metrics for measuring alpha-diversity (Chao *et al*. 2012, 2014) since most existing diversity metrics such as species richness, Shannon entropy, and Simpson index turned out to be special cases (or functions) of the Hill numbers. To take advantages of the Hill numbers as general biodiversity metrics, Ma (2018a) extended the classic SAR to more general diversity-area relationship (DAR) by substituting the species richness with general diversity measured with Hill numbers. In the present study, we applied the DAR approach to address the previous raised questions regarding the changes of DT microbiome diversity, both across individuals (inter-individual) and across DT sites (intra-individual or intra-DT).

To investigate the inter-individual and intra-individual (intra-DT) diversity scaling patterns with the DAR approach, we used a dataset originally collected by Segata *et al* (2012). Their study collected ten microbiome samples from each of over 200 individuals' digestive tract [buccal mucosa (BM), keratinized gingiva (KG), hard palate (HP), throat (Th), palatine tonsils (PT),

tongue dorsum (TD) and saliva (Sal), supraginval (SupP), subgingival plaques (SubP), and stool (Stool)]. The dataset provides an ideal opportunity for us achieve the objective of this study—analyzing the inter-individual and intra-individual diversity scaling of the human DT microbiomes.

Materials and Methods

A brief description of the DT (digestive tract) microbiome dataset

The DT microbiome dataset we reanalyzed in this study was first reported by Segata et al. (2012), which is part of the Human Microbiome Project (HMP). Total 2078 DT microbiome samples were collected from 242 healthy adults aged from 18 to 40 years old, who were enrolled in the HMP. The ten DT sites sampled included seven from oral cavity (BM, KG, HP, Th, PT, TD and Sal), two from oropharynx (SupP and SubP), and one from gut (Stool). The operational taxonomic unit (OTU) tables and the metadata information on the individuals are available at https://www.hmpdacc.org/, and more detailed information on the dataset is referred to Segata et al. (2012).

Computational procedures for the DAR analysis

Definitions of alpha diversities

We applied the Hill numbers (Hill 1973, Chao et al. 2012, 2014) to measuring alpha diversity, which are defined as:

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

where D is the diversity Hill numbers, q (=0, 1, 2, …) is the order number of diversity, S is number of species (or OTUs), and p_i is the relative abundance of OTU i. When q = 1, the Hill number is not defined, but we can figure out its limit as q approaches to 1 as follows:

$$ {}^{1}D = \lim_{q \to 1} {}^{q}D = \exp\left(-\sum_{i=1}^{S} p_i \log(p_1) \right) \tag{2} $$

The Hill numbers are a series of diversity measures corresponding different diversity orders (q), where q determines the weight of relative frequencies of species abundances. When q = 0, species abundance is not involved in the calculation, and ${}^{0}D$ is the number of OTUs or the species richness. When q = 1, ${}^{1}D$ equals the *exponential* of Shannon entropy, which represents the number of typical or common species in the community. When q = 2, ${}^{2}D$ is equal to the

reciprocal of Simpson index and represents the number of species with high abundance. Generally, ${}^{q}D$ represents the diversity of a community with $x = {}^{q}D$ equally abundant species.

DAR analysis

According to Ma (2018a), we selected and used two DAR models for the DT microbiome in this study, in which one adopted power law (PL) model, and another adopted power law with exponential cutoff (PLEC) model. The PL model is:

$$ {}^{q}D = cA^{z} \tag{3}$$

where ${}^{q}D$ is diversity measured in the q-th order Hill numbers, A is *area*, and c & z are the PL parameters.

The PLEC model is:

$$ {}^{q}D = cA^{z}\exp(dA) \tag{4}$$

where d is a third parameter with taper-off effect, and $\exp(dA)$ is the exponential decay term that eventually overwhelms the power law behavior when A becomes very large.

We transformed the non-linear Equations (3)(4) into log-linear regression equations (5)(6) to estimate the parameters of PL and PLEC models, respectively:

$$ \ln(D) = \ln(c) + z\ln(A) \tag{5}$$

$$ \ln(D) = \ln(c) + z\ln(A) + dA \tag{6}$$

Four important DAR parameters and corresponding profiles

According to Ma (2018a) and Ma & Li (2018), there are four most important DAR parameters, including the diversity scaling parameter (z), pair-wise diversity overlap (PDO or g), maximal accrual diversity (MAD or $D_{\max}$), and ratio of individual diversity to population accural diversity (RIP).

(*i*) As the slope or tangent of the PL-DAR model (Eqns. 3 & 5), the z-value was termed as the diversity scaling parameter.

(*ii*) The PDO or g was defined as,

$$ g = 2 - 2^{z} \tag{7}$$

where z is the scaling parameter of the PL-DAR model. The range of g is generally between 0 and 1. If $z = 1$, then $g = 0$ and there is no overlap or similarity. If $z = 0$, then $g = 1$ and there is a total overlap or similarity.

(*iii*) The MAD or $D_{\max}$ was defined based on the PLEC-DAR model (Eqns. 4 & 6), that is,

$$ {}^{q}D_{\max} = c\left(-\frac{z}{d}\right)^{z}\exp(-z) = cA_{\max}^{z}\exp(-z) \tag{8} $$

where $A_{\max} = -z/d$ is the number of individuals (microbiome samples) needed to reach the MAD, and c & d are parameters of the PLEC-DAR model.

(*iv*) The RIP was defined as,

$$ {}^{q}RIP = {}^{q}c/{}^{2}D_{\max} \tag{9} $$

where ${}^{q}c$ is the parameter of PL-DAR model at diversity order of q, and ${}^{q}D_{\max}$ is the MAD that can be computed with Equation (8).

Ma (2018) and Ma & Li (2019) also defined the relationships between these four parameters and the diversity order (q) as the DAR profile, PDO profile, MAD profile and RIP profile, respectively.

Design for DAR analysis

Our analysis consists of two parts, inter-individual (cross-individual) DAR analysis and intra-individual (cross-DT site) DAR analysis. Based on the inter-individual DAR analysis, we can investigate diversity scaling for each 10 DT sites cross over 200 individuals. We built PL-DAR and PLEC-DAR models for each DT sites, and further conduct permutation test (randomization test) for the DAR parameters (i.e., z and $D_{\max}$) of different DT sites. The procedure for randomization test is referred to Collingridge (2013), in which the number of permutations or re-samplings was set to 1000 times. The p-value of randomization test can be used to determine the significance of differences. It is noted that parameter c of the PL model indicates the diversity in the first unit of area to accrue. Thus, to exclude the influence of the accrual order of area unit on parameter c, we randomly permutated the area units to be accumulated each time the DAR model was built. In the inter-individual DAR analysis, we repeated this re-sampling procedure 100 times, and adopted the averages of the model parameters from the 100 times of DAR fittings as

the final model parameters of the inter-individual DAR model for the DT site under investigation. The detailed computational procedure can be referred to Ma (2018a, 2018b). Based on the intra-individual DAR analysis, we can investigate intra-DT diversity scaling cross 10 DT sites. The steps of intra-individual DAR analysis are as follow: (*i*) We first selected a random sample from all samples of each 10 DT sites, and treated the set of these 10 random samples as an "individual". (*ii*) We built PL-DAR and PLEC-DAR models for the "individual", in which area units (DT sites) were accumulated in order of anatomy structure from the oral cavity to the intestinal tract. (*iii*) We repeated steps (*i*)-(*ii*) for 1000 times, and adopted the averages of the model parameters from the 1000 times of DAR fittings as the final model parameters of the intra-individual DAR model. In addition, beside genus taxon level, we also analyzed the DAR patterns of other taxa including phylum, class, order, and family taxa.

Results

Inter-individual DAR for each DT site

At each of 5 taxon levels, we built PL-DAR and PLEC-DAR models for each of the 10 sites of human DT microbiome. Table 1 listed the results of fitting DAR models for each 10 DT sites at the genus taxon level, including the diversity order (q) of the Hill numbers, the mean model parameters (z, c, d, g, $D_{\max}$ and RIP) and measures for goodness-of-fitting (R and p-value). N is the number of successful fittings out of 100 re-samplings, as explained previously. Tables S1-S4 listed the results at other taxon levels, i.e., phylum, class, order, and family. Tables 2A & 2B listed results from the permutation tests for the differences in scaling parameter (z) and $D_{\max}$ of different DT microbiome sites. From these results, we summarize the following findings:

(*i*) *DAR profile*: At genus level, the average scaling parameter (z) of the 10 DT sites across diversity order q = 0-3 is z = (0.294, 0.038, 0.020, 0.014), and their standard errors ranged from 0.003 to 0.008 (as shown in Table 1 & Figure 1). For each DT microbiome site, the z-values monotonically decreased with the diversity order q. As shown in Tables S1-S4, the scaling parameter (z) gradually decreased with the taxon level, and for example, the average scaling parameter (z) at phylum level across q = 0-3 is z = (0.108, 0.013, 0.009, 0.007). As shown in Table 2A, no significant differences in scaling parameter (z) at species-level were detected among 10 DT sites, and the detailed results of randomization tests were listed in Table S5.

(*ii*) *PDO profile*: PDO or parameter g characterizes the overlap or similarity between pair-wise microbiomes. As shown in Table 1, at genus level, the average PDO parameter (g) of the 10 DT sites across diversity order q = 0-3 is g = (0.773, 0.973, 0.985, 0.989), and their standard errors ranged from 0.002 to 0.007. In contrast to the diversity scaling parameter (z), the PDO parameter (g) increased with either diversity order q or taxon level (see Tables S1-S4).

(*iii*) *MAD profile*: As shown in Table 1, at genus level, the average D_{max} of the 10 DT sites across diversity order q = 0-3 is D_{max} = (288.7, 14.2, 8.2, 6.6). MAD or parameter D_{max} can be considered as a proxy of potential or "dark" diversity, which can be used to estimate microbial biodiversity of a DT sites for a human population. For example, at taxonomic genus, the maximal accrual of species richness (Hill numbers for q = 0) across individuals is around 289. Similar to the DAR profile, the MAD profile of each DT sites decreased with diversity order q (Table 1 & Figure 1). As shown in Tables S1-S4, the D_{max} decreased with the taxon level, and for example, the average D_{max} at phylum level across q = 0-3 is D_{max} = (14.7, 4.0, 3.3, 3.1). We test the difference in parameter D_{max} between each pair of DT sites by using randomization test, and results were listed in Table 2B. At diversity order q = 0, 2 out of 45 or 4.4% comparisons between 10 DT sites exhibited statistically significant differences. These two comparisons with difference were SAL *vs*. Stool and SupP *vs*. Stool. At diversity order q = 1-3, there were 73.3% (33/45), 57.8% (26/45) and 68.9% (31/45) comparisons with significant differences, respectively. Please see Table S5 for the detailed results of randomization tests.

(*iv*) *RIP profile*: As shown in Table 1, at genus level, the average RIP of the 10 DT sites across diversity order q = 0-3 is *RIP* = (19.1, 83.1, 90.8, 93.4), and their standard errors ranged from 1.0 to 3.3. The RIP profiles of each DT sites monotonically increased with q (Table 1 & Figure 1). RIP can characterize the relationship between individual-level diversity and population-level diversity. For example, at diversity order q = 0, RIP = 19.1 indicate that an average individual can represent for approximately 19% of population diversity.

250 **Table 1**. The DAR (Diversity-Area Relationship) models for each of the 10 DT (digestive tract) microbiome sites at the *genus* taxon-level (see Supplementary Tables S1-S4 for the results at other taxon levels)

Site	Diversity Order	Power Law (PL)						PL with Exponential Cutoff (PLEC)								RIP
		z	*c*	*R*	*p*-value	*g*	**N*	*z*	*d*	*c*	*R*	*p*-value	**N*	A_{max}	D_{max}	
BM	*q=0*	0.289	59.145	0.984	0.000	0.778	100	0.349	-0.001	51.009	0.992	0.000	100	323	270.1	21.9
	q=1	0.033	6.475	0.538	0.044	0.976	88	0.058	0.000	6.086	0.675	0.003	99	130	7.6	85.2
	q=2	0.009	3.105	0.526	0.029	0.993	92	0.011	0.000	3.089	0.670	0.006	95	396	3.3	94.1
	q=3	0.003	2.482	0.521	0.019	0.997	94	0.000	0.000	2.502	0.669	0.007	99	2	2.5	99.3
HP	*q=0*	0.302	64.586	0.99	0.000	0.767	100	0.343	-0.001	58.323	0.995	0.000	100	453	337.7	19.1
	q=1	0.018	9.564	0.445	0.082	0.987	83	0.034	0.000	9.198	0.613	0.007	95	117	10.5	91.1
	q=2	-0.015	4.545	0.474	0.055	1.009	84	-0.025	0.000	4.655	0.620	0.010	95	141	4.2	100**
	q=3	-0.021	3.449	0.489	0.054	1.014	91	-0.038	0.000	3.600	0.629	0.013	98	121	3.1	100**
KG	*q=0*	0.345	36.416	0.983	0.000	0.73	100	0.407	-0.001	31.187	0.990	0.000	100	364	229.3	15.9
	q=1	0.049	5.675	0.56	0.037	0.964	93	0.096	-0.001	5.058	0.692	0.003	98	115	7.2	78.8
	q=2	0.011	3.254	0.516	0.028	0.991	88	0.027	0.000	3.124	0.660	0.006	96	93	3.4	95.7
	q=3	0.001	2.649	0.5	0.106	0.998	80	0.007	0.000	2.612	0.654	0.014	97	67	2.7	98.1
SubP	*q=0*	0.280	56.997	0.991	0.000	0.786	100	0.303	0.000	53.839	0.995	0.000	100	724	292.5	19.5
	q=1	0.038	17.567	0.709	0.008	0.973	98	0.075	-0.001	16.023	0.817	0.000	100	113	21.2	82.9
	q=2	0.046	11.635	0.686	0.018	0.967	96	0.092	-0.001	10.381	0.810	0.008	99	112	14.6	79.7
	q=3	0.054	9.621	0.68	0.015	0.962	94	0.104	-0.001	8.491	0.801	0.000	100	114	12.5	77.0
SupP	*q=0*	0.281	48.327	0.99	0.000	0.785	100	0.288	0.000	47.418	0.993	0.000	100	2164	325.4	14.9
	q=1	0.038	15.364	0.598	0.045	0.973	93	0.072	-0.001	14.098	0.736	0.000	100	118	18.5	83.0
	q=2	0.043	10.454	0.583	0.024	0.969	91	0.080	-0.001	9.535	0.720	0.001	99	123	12.9	81.0
	q=3	0.046	8.793	0.567	0.031	0.967	90	0.083	-0.001	8.029	0.711	0.004	99	128	11	79.9
SAL	*q=0*	0.293	61.992	0.99	0.000	0.775	100	0.324	-0.001	57.512	0.994	0.000	100	536	319.3	19.4
	q=1	0.027	16.827	0.604	0.019	0.981	94	0.055	-0.001	15.721	0.745	0.007	97	101	19.2	87.6
	q=2	0.023	10.486	0.534	0.033	0.983	90	0.046	0.000	9.924	0.685	0.003	99	105	11.8	88.9
	q=3	0.022	8.628	0.519	0.034	0.984	91	0.043	0.000	8.207	0.669	0.000	100	107	9.6	89.9
TD	*q=0*	0.274	44.835	0.988	0.000	0.791	100	0.291	0.000	42.905	0.992	0.000	100	945	235.9	19.0
	q=1	0.036	12.833	0.595	0.032	0.975	92	0.071	-0.001	11.752	0.732	0.000	100	115	15.3	83.9
	q=2	0.036	8.593	0.507	0.044	0.974	87	0.070	-0.001	7.893	0.661	0.003	99	117	10.3	83.4
	q=3	0.031	7.199	0.479	0.038	0.978	89	0.062	-0.001	6.659	0.636	0.013	98	113	8.4	85.7
PT	*q=0*	0.291	56.599	0.986	0.000	0.776	100	0.329	-0.001	51.522	0.991	0.000	100	492	284.6	19.9
	q=1	0.045	12.897	0.542	0.025	0.968	90	0.087	-0.001	11.612	0.678	0.000	100	117	16.1	80.1
	q=2	0.027	7.546	0.456	0.050	0.980	86	0.062	-0.001	6.910	0.583	0.019	95	100	8.7	86.7
	q=3	0.012	6.007	0.467	0.047	0.990	89	0.037	0.000	5.641	0.588	0.021	96	82	6.4	93.9
Th	*q=0*	0.330	60.947	0.989	0.000	0.743	100	0.369	-0.001	55.313	0.993	0.000	100	513	382	16.0
	q=1	0.042	13.585	0.551	0.064	0.970	86	0.081	-0.001	12.342	0.701	0.005	98	113	16.7	81.4
	q=2	0.017	7.815	0.477	0.062	0.987	86	0.043	0.000	7.330	0.632	0.015	94	89	8.5	91.9
	q=3	0.003	6.135	0.484	0.035	0.997	91	0.020	0.000	5.894	0.639	0.006	99	65	6.3	97.4
Stool	*q=0*	0.259	54.055	0.976	0.000	0.803	100	0.323	-0.001	45.925	0.986	0.000	100	303	210.5	25.7
	q=1	0.051	7.822	0.51	0.056	0.963	86	0.107	-0.001	6.780	0.691	0.004	99	115	10.1	77.5
	q=2	0.002	4.116	0.461	0.066	0.997	88	0.028	0.000	3.854	0.636	0.018	97	65	4.2	98.0
	q=3	-0.010	3.245	0.46	0.027	1.006	84	0.002	0.000	3.146	0.629	0.012	97	9	3.2	100**
q=0	**Mean**	0.294	54.390	0.987	0.000	0.773	100	0.333	-0.001	49.495	0.992	0.000	100	682	288.7	19.1
	Std Error	0.008	2.769	0.001	0.000	0.007	0	0.012	0.000	2.576	0.001	0.000	0	176	17.1	1.0
q=1	**Mean**	0.038	11.861	0.565	0.041	0.973	90	0.074	-0.001	10.867	0.708	0.003	99	115	14.2	83.1
	Std Error	0.003	1.347	0.022	0.007	0.002	1	0.007	0.000	1.248	0.017	0.001	1	2	1.6	1.3
q=2	**Mean**	0.020	7.155	0.522	0.041	0.985	89	0.043	0.000	6.670	0.668	0.009	97	134	8.2	90.8
	Std Error	0.006	1.013	0.022	0.005	0.004	1	0.011	0.000	0.894	0.020	0.002	1	30	1.3	2.8

q=3	**Mean**	0.014	5.821	0.517	0.041	0.989	89	0.032	0.000	5.478	0.663	0.009	98	81	6.6	93.4
	Std Error	0.008	0.861	0.021	0.008	0.005	1	0.014	0.000	0.750	0.018	0.002	0	14	1.2	3.3

* The model parameters were computed from 100 times of repeated DAR modeling based on 100 times of re-sampling; *N* is the number of successful DAR model-fitting out of 100 times of re-sampling.

255 ** The theoretical maximum of RIP is 100%, any RIP exceeding 100% may be caused by estimation errors.

Table 2A. Summary of the randomization tests in Table S5: the pair-wise comparisons of the DAR parameter (z) between the ten sites of the human digestive microbiome (Genus level), for

260 each diversity order q.

Digestive Sites	**BM**	**HP**	**KG**	**PT**	**SAL**	**Stool**	**SubP**	**SupP**	**TD**	**Th**
BM	NA	0000	0000	0000	0000	0000	0000	0000	0000	0000
HP	0000	NA	0000	0000	0000	0000	0000	0000	0000	0000
KG	0000	0000	NA	0000	0000	0000	0000	0000	0000	0000
fPT	0000	0000	0000	NA	0000	0000	0000	0000	0000	0000
SAL	0000	0000	0000	0000	NA	0000	0000	0000	0000	0000
Stool	0000	0000	0000	0000	0000	NA	0000	0000	0000	0000
SubP	0000	0000	0000	0000	0000	0000	NA	0000	0000	0000
SupP	0000	0000	0000	0000	0000	0000	0000	NA	0000	0000
TD	0000	0000	0000	0000	0000	0000	0000	0000	NA	0000
Th	0000	0000	0000	0000	0000	0000	0000	0000	0000	NA
Significance (%)	0	0	0	0	0	0	0	0	0	0

Table 2B. Summary of the randomization tests in Table S5: the pair-wise comparisons of the DAR parameter (D_{max}) between the ten sites of the human digestive microbiome (Genus level) for each diversity order q.

Digestive Sites	**BM**	**HP**	**KG**	**PT**	**SAL**	**Stool**	**SubP**	**SupP**	**TD**	**Th**
BM	NA	0100	0000	0101	0101	0000	0101	0101	0101	0101
HP	0100	NA	0100	0111	0111	0000	0111	0111	0111	0111
KG	0000	0100	NA	0111	0111	0000	0111	0111	0111	0111
PT	0101	0111	0111	NA	0111	0100	0111	0011	0000	0000
SAL	0101	0111	0111	0111	NA	1111	0111	0000	0100	0011
Stool	0000	0000	0000	0100	1111	NA	0111	1111	0111	0100
SubP	0101	0111	0111	0111	0111	0111	NA	0110	0111	0111
SupP	0101	0111	0111	0011	0000	1111	0110	NA	0111	0011
TD	0101	0111	0111	0000	0100	0111	0111	0111	NA	0000
Th	0101	0111	0111	0000	0011	0100	0111	0011	0000	NA
***q*=0 (%)**	0	0	0	0	11.1	22.2	0	11.1	0	0
***q*=1 (%)**	77.8	88.9	77.8	66.7	77.8	66.7	100	66.7	77.8	55.6
***q*=2 (%)**	0	66.7	66.7	55.6	66.7	44.4	88.9	77.8	55.6	55.6
***q*=3 (%)**	66.7	66.7	66.7	66.7	77.8	44.4	88.9	77.8	66.7	66.7

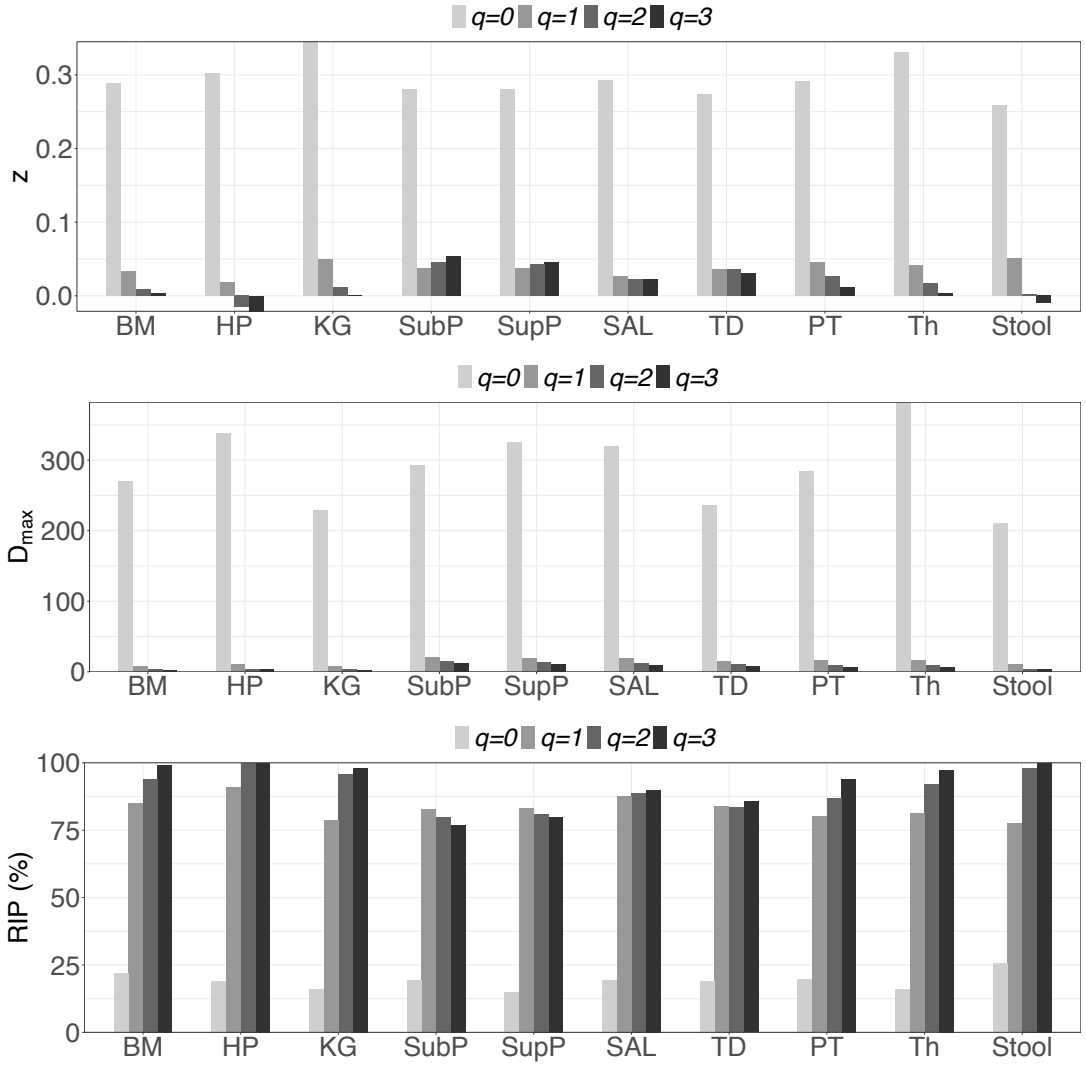

Fig 1. Graphs of three important profiles from inter-individual DAR models for each of 10 DT microbiome sites at genus taxon level, including DAR profiles (*z*-*q* patterns), MAD profiles (D_{max}-*q* patterns), and RIP profiles (*RIP*-*q* patterns). Bar color indicates diversity order. The *x*-axis shows the DT sites: buccal mucosa (BM), keratinized gingiva (KG), hard palate (HP), throat (Th), palatine tonsils (PT), tongue dorsum (TD) and saliva (Sal), supraginval (SupP), subgingival plaques (SubP), and stool (Stool).

Intra-DT diversity scaling cross DT sites with intra-individual DAR models

At each taxon level, we built the PL- and PLEC-DAR models cross 10 DT sites to investigate intra-DT distribution pattern. Table 3 listed the results of fitting intra-DAR models at all five taxon levels, including the same parameters as Table 1. *N* is the number of successful fittings out of 1000 re-samplings, as explained in section of materials and methods. When $q = 0$, fitting to both intra-DAR models were failed at phylum level. The reason is that different DT sites had the same number of phyla in 890 out of 1000 re-samplings, for which result in the relationship between ln(*D*) and ln(*c*) being equivalent to a line parallel to the *x*-axis, and the estimation of

goodness-of-fitting R and p-value being failed. From Table 3, we summarize the following findings:

(*i*) *DAR profile*: At genus level, the scaling parameter (z) across diversity order q = 0-3 is z = (0.417, 0.512, 0.535, 0.493). DAR profile increased with q at q = 0-3, but slightly decreased at q = 4. Compared with inter-individual DAR profile, there were not much differences in z-values between at q = 0 and other diversity orders in intra-individual diversity scaling. Similar to the inter-individual diversity scaling, the scaling parameter (z) also decreased with the taxon level, but dropped relatively slowly.

(*ii*) *PDO profile*: The trends of PDO profiles over diversity order q were contrary to those of DAR profiles.

(*iii*) *MAD profile*: At genus level, D_{max} across diversity order q = 0-3 is D_{max} = (140.0, 18.2, 12.4, 5.8). MAD or parameter D_{max} offers estimates for the potential microbial diversity in the whole human DT. For example, at taxonomic genus, the theoretical maximal accrual of species richness (Hill numbers for q = 0) across all DT sites is around 140. Similar to the inter-individual MAD profile, D_{max} decreased with diversity order q and taxon level.

(*iv*) *RIP profile*: At genus level, RIP across diversity order q = 0-3 is *RIP* = (32.5, 28.7, 20.8, 36.1). The RIP profiles of each DT sites monotonically increased with q. Compared with inter-individual DAR models, the trends of RIP over either diversity order q or taxon level were less obvious in intra-individual DAR models.

305 **Table 3**. The intra-DAR (intra-individual Diversity-Area Relationship) models

Taxon	Diversity Order	Power Law (PL)						PL with Exponential Cutoff (PLEC)								RIP
		z	*c*	*R*	*p*-value	*g*	*N	*z*	*d*	*c*	*R*	*p*-value	A_{max}	D_{max}	*N	
Genus	*q=0*	0.417	45.468	0.962	0.000	0.665	1000	0.555	-0.030	44.168	0.974	0.000	41	140.0	454	32.5
	q=1	0.512	5.228	0.900	0.003	0.574	933	0.439	0.013	5.490	0.929	0.004	22	18.2	584	28.7
	q=2	0.535	2.583	0.886	0.004	0.551	865	0.273	0.051	2.915	0.920	0.005	29	12.4	564	20.8
	q=3	0.493	2.094	0.880	0.004	0.593	838	0.172	0.063	2.401	0.915	0.006	15	5.8	565	36.1
Family	*q=0*	0.261	31.785	0.959	0.000	0.802	1000	0.391	-0.030	30.938	0.975	0.000	34	60.6	604	52.5
	q=1	0.472	4.740	0.906	0.002	0.613	923	0.467	-0.004	4.988	0.935	0.003	22	14.3	580	33.1
	q=2	0.526	2.581	0.892	0.003	0.560	885	0.338	0.034	2.954	0.925	0.004	28	9.5	542	27.2
	q=3	0.498	2.119	0.886	0.004	0.588	859	0.226	0.050	2.497	0.918	0.005	21	6.7	530	31.6
Order	*q=0*	0.292	16.395	0.954	0.000	0.776	1000	0.425	-0.032	16.248	0.973	0.000	53	34.5	596	47.5
	q=1	0.347	3.869	0.892	0.003	0.728	880	0.461	-0.030	3.912	0.927	0.005	23	8.9	604	43.5
	q=2	0.422	2.430	0.888	0.003	0.660	850	0.393	-0.001	2.588	0.920	0.005	22	6.4	576	38.0
	q=3	0.427	2.040	0.887	0.003	0.656	839	0.303	0.020	2.219	0.919	0.005	17	5.3	547	38.5
Class	*q=0*	0.222	12.756	0.935	0.000	0.834	999	0.335	-0.028	12.591	0.962	0.001	26	21.6	660	59.1
	q=1	0.383	3.133	0.892	0.003	0.696	918	0.492	-0.027	3.133	0.932	0.004	25	7.8	614	40.2
	q=2	0.455	2.022	0.891	0.003	0.629	894	0.416	0.003	2.153	0.924	0.005	27	6.0	542	33.7
	q=3	0.451	1.738	0.888	0.004	0.633	891	0.322	0.023	1.895	0.923	0.005	24	4.9	514	35.5
Phylum	*q=0*	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	q=1	0.285	2.366	0.880	0.003	0.782	900	0.380	-0.025	2.382	0.916	0.005	18	4.5	617	52.6
	q=2	0.319	1.749	0.875	0.004	0.753	877	0.399	-0.024	1.820	0.909	0.006	49	3.8	579	46.0
	q=3	0.311	1.576	0.871	0.004	0.759	866	0.374	-0.020	1.647	0.903	0.007	25	3.3	561	47.8

* The model parameters were computed from 1000 times of repeated DAR modeling based on 1000 times of re-sampling; *N* is the number of successful DAR model-fitting out of 1000 times of re-sampling.

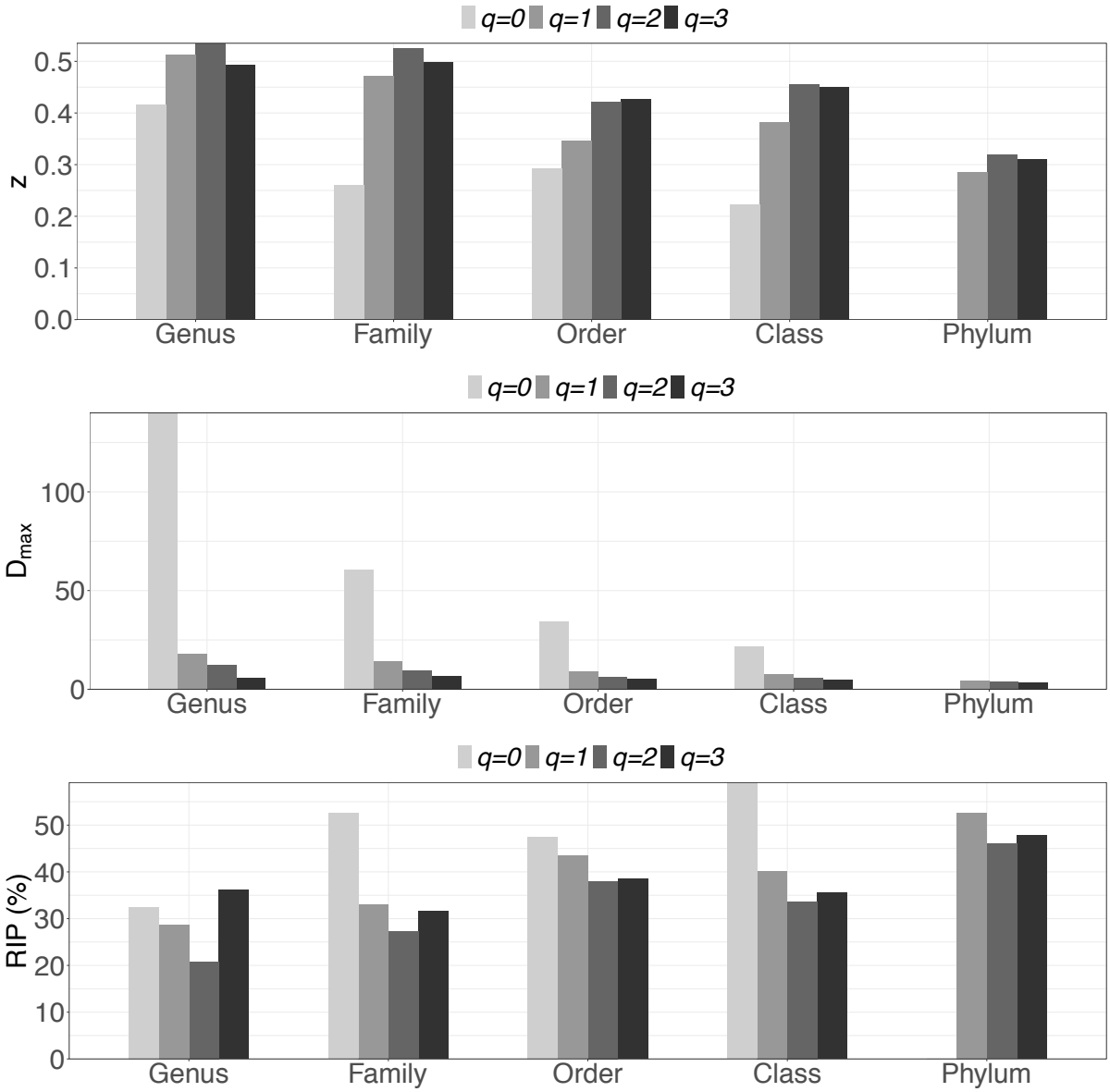

Figure 2. Graphs of three important profiles from intra-individual DAR models for each of 5 taxon levels, including DAR profiles (*z*-*q* patterns), MAD profiles (D_{max}-*q* patterns), and RIP profiles (*RIP*-*q* patterns). Bar color indicates diversity order.

Conclusions and Discussion

Understanding the biogeography or the spatial distribution of biodiversity is of critical significance both theoretically and practically. Theoretically, biogeography shows a big picture of community/metacommunity patterns on a larger scale, in our case, the inter-individual microbiome distribution in a human population cohort. Practically, biogeography of human microbiome is of obvious importance for public health and personified medicine. For instance, understanding the inter-individual heterogeneity is essential for studying and implementing

microbiome intervention treatments such as fecal transplantation for treating certain microbiome-associated diseases (Weingarden & Vaughn 2017, Kump et al. 2013, Wei et al. 2015, Ishikawa et al. 2017, 2018, Cohen et al. 2019). Similarly, studying the intra-individual (along the human DT in this study) diversity scaling is also of obvious significance. We apply the unified DAR approach to investigate both inter-individual and intra-DT (intra-individual) diversity scaling (across individual and across DT sites) of the human DT microbiomes with Segata et al. (2012) datasets of 10 DT sites sampled from more than 200 healthy individuals.

In terms of the inter-individual microbial diversity scaling (Table 1 & Table 2A and 2B), the diversity scaling parameter (z) of all 10 DT sites seems invariant with the site—the z-values of all sites did not show significant statistical differences. At genus level, the average scaling parameter (rate) (z) across 10 sites is 0.294 at diversity order q=0 (species richness), 0.038 at q=1 (Shannon entropy), 0.020 at q=2 (Simpson index), and 0.014 at q=3. The scaling rate of *species richness* (Hill number for q=0) is nearly 10 times larger than those of other diversity orders, *e.g.,* community evenness measured with Shannon entropy. These results suggest that the inter-individual differences in DT microbiome diversity are primarily in the number of microbial genus —species richness, as demonstrated by much higher scaling rate, rather than in general community diversity as demonstrated by Hill numbers for q>1. The inter-individual diversity scaling parameters (z) obtained in this study is also consistent with previous study by Ma (Citations: 2018a, Ecology and Evolution; 2018b, Microbial Ecology), in which single gut microbiome diversity scaling was investigated.

Besides inter-individual diversity scaling parameter (z), another DAR parameter RIP (ratio of individual to population level diversity) also revealed that the critical differences between individuals lie in species richness (q=0), rather than in community diversity (q>0). The average RIP for species richness (q=0) of 10 DT sites is 19.1% with standard error of 1.0 only, while RIP for general community diversity (q=1 to 3) ranged from 83.1% to 93.4%. These RIP numbers indicates that an average individual can host approximately 20% of microbial genus owned by a whole population, while the microbial diversity of an individual may exceed 90% the total diversity of a population from which the individual comes from. Furthermore, all 10 DT exhibited very similar inter-individual diversity scaling as described above, which is evidenced by the rather small standard error of the average z and RIP across the 10 DT sites. To the best of our knowledge, no previous studies have addressed the RIP of gut microbiomes.

We also systematically investigated the inter-individual diversity scaling on other four taxa including phylum, class, order, and family (Table S1-S4). The scaling patterns are similarly to the previously summarized genus-level scaling, but the diversity scaling parameter (z) generally decreases with the taxon level. That is, the inter-individual differences in diversity decreases with the higher taxonomic orders. This should be expected since higher diversity order such as phyla and classes are more general (rough) classifications and the similarity in diversity should certainly be higher at more general taxonomic scales. Higher similarity in diversity is equivalent to lower diversity scaling parameter (z), *i.e.,* slower scaling rate. To the best of knowledge, this study should be the first one that studies diversity scaling at taxonomic level beyond species.

In terms of intra-DT diversity scaling patterns (Table 3), at the genus level, the diversity scaling parameter (z) at various diversity orders (q=0-3) is actually more similar that inter-individual scaling, which is indicated by the relatively narrow range of z-values {0.417 (q=0), 0.512 (q=1), 0.535 (q=2), 0.493 (q=3)}. Similarly, the RIP vector {32.5% (q=0), 28.7% (q=1), 20.8% (q=2), 36.1 (q=3)} also exhibited a rather range, compared with previous inter-individual. These findings suggest that intra-DT heterogeneity seems to be universally stronger than the inter-individual heterogeneity. This is obviously determined by the human biology, since for intra-DT diversity scaling, we are comparing "apples" and "oranges" (*e.g*., oral site *vs*. gut), while inter-individual diversity scaling was comparing the "apples from two apples trees". Therefore, comparing them may not be that meaningful. The important insight our study revealed is that (*i*) each DT site hosts approximately 1/5 to 1/3 of the whole DT diversity, and (*ii*) there should be significantly overlaps (similarity) among the DT sites as inferred by (*i*). This high similarity in intra-DT diversity can be explained by the biological fact that DT is a continuum in which microbial dispersal occurs routinely. On the other hand, the heterogeneity in diversity scaling can be explained that the DT continuum is not homogenous either. In fact, the DT is differentiated as four different niches hosting some functionally different microbial species as revealed in Segata *et al*. (2012) original study, upon the datasets of which our study is based.

Similar to the inter-individual diversity scaling, we also investigated the intra-DT diversity scaling on other four taxon levels (phylum, class, order, and family) (Table 3) beyond genus level. The pattern is similar to previous inter-individual diversity scaling. That is, at higher taxonomic level, the difference becomes smaller or the similarity become larger, perhaps like

using telescope to observe remote landscapes. For example, at the taxonomic *class* level, the RIP for q=0 was 59.1%, suggesting that an average individual can host approximately 60% of microbial classes of a whole population owns.

A minor limitation of this study is that the species-level DAR analysis was missing given that the original raw sequencing reads reported in Segata *et al*. (2012) were only binned to genus and above taxon levels. Since in many cases, the species level (or 97% similarity level) OTUs are simply a number appended to genus name, and may be of limited biomedical significances. In the meantime, the annotations at higher taxon levels (genus, family, order, class, and phylum) should be rather stable, and the analyses of their diversity scaling can be more useful practically. In fact, to the best of our knowledge, this study should be the first comprehensive analysis of DT microbiomes at and above genus levels.

Data availability

The datasets are available from the online supplementary material of "Segata *et al* (2012) Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol.* 13(6):R42."

Funding

This research is supported by Open Grants from the State Key Laboratory of Genetic Resources and Evolution (No. (#GREKF18-14, GREKF19-07) and the Yunnan Province Local University (Part) Basic Research for Youths (No.2019FH001-117 and No. 2018FD086) and the Yunnan provincial department of education scientific research fund project (No.2019J0646 and No.2017ZDX076).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

HJC & BY conducted the data analysis; XX and QL participated in the results interpretations and discussion; LD and ZSM designed the study; HJC, LD and ZSM wrote and revised the manuscript; all authors approved the submission.

References

Chao A, Chiu CH, Hsieh TC (2012) Proposing a resolution to debates on diversity partitioning. *Ecology*, 93(9): 2037-2051.

Chao A, Chiu CH, & Jost L (2014) Unifying species diversity, phylogenetic diversity, functional diversity and related similarity and differentiation measures through Hill numbers. *Annual Reviews of Ecology, Evolution, and Systematics*, 45: 297–324.

Chao A, Gotelli NG, Hsieh TC, et al. (2014) Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species biodiversity studies. *Ecological Monographs*, 84, 45–67.

Cohen LJ, Cho JH, Gevers D, Chu H (2019) Genetic Factors and the Intestinal Microbiome Guide Development of Microbe-Based Therapies for Inflammatory Bowel Diseases. Gastroenterology., 156(8):2174-2189. doi: 10.1053/j.gastro.2019.03.017.

Connor EF & McCoy ED (1979) The statistics and biology of the species–area relationship. *The American Naturalist*, 113: 791–833.

Drakare S, Lennon JJ & Hillebrand H (2006) The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecology Letters*, 9: 215–227

Harte J, Smith AB, and Storch D (2009) Biodiversity scales from plots to biomes with a universal species-area curve. *Ecology Letters,* 12: 789–797.

He FL & Hubbell SP (2011) Species–area relationships always overestimate extinction rates from habitat loss. *Nature,* 473: 368-371

Helmus MR, Mahler DL & Losos JB (2014) Island biogeography of the Anthropocene. *Nature,* 513:543-547

Hill MO (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology,* 54:427-342.

HMP Consortium (2012a) A framework for human microbiome research. *Nature*, 486(7402):215.

HMP Consortium (2012b) Structure, function and diversity of the healthy human microbiome. *Nature*, vol; 486: 207-214. doi:10.1038/nature11234

Ishikawa D, Sasaki T, Osada T, et al. (2017) Changes in Intestinal Microbiota Following Combination Therapy with Fecal Microbial Transplantation and Antibiotics for Ulcerative Colitis. Inflamm Bowel Dis., 23(1):116-125. doi: 10.1097/MIB.0000000000000975.

Ishikawa D, Sasaki T, Takahashi M, et al. (2018) The Microbial Composition of Bacteroidetes Species in Ulcerative Colitis Is Effectively Improved by Combination Therapy With Fecal Microbiota Transplantation and Antibiotics. *Inflamm Bowel Dis*., 24(12):2590-2598. doi: 10.1093/ibd/izy266.

Jost L (2007) Partitioningdiversityintoindependentalphaandbetacomponents. *Ecology,* 88: 2427–2439. doi: 10.1890/06-1736.1

Kump PK, Gröchenig HP, Lackner S, et al. (2013) Alteration of intestinal dysbiosis by fecal microbiota transplantation does not induce remission in patients with chronic active ulcerative colitis. Inflamm Bowel Dis., 19(10): 2155-2165. doi: 10.1097/MIB.0b013e31829ea325.

Li W & Ma ZS (2019) Diversity scaling of human vaginal microbial communities. *Zoological Research*, 40(6): 587-594 doi: 10.24272/j.issn.2095-8137.2019.068

Lomolino MV (2000) Ecology's most general, yet protean pattern: the species–area relationship. *Journal of Biogeography*, 27: 17–26.

Lozupone CA, Stombaugh JI, Gordon J, Jansson JK, Knight R (2012) Diversity, stability and resilience of the human gut microbiota. *Nature*, 489: 220-230.

Ma ZS (2017). Measuring microbiome diversity and similarity with hill numbers. *Metagenomics*, 2018: 157–178 doi: 10.1016/B978-0-08-102268-9.00008-2

Ma ZS (2018a) DAR (diversity-area relationship): Extending classic SAR (species-area relationship) for biodiversity and biogeography analyses. *Ecology and Evolution*, 8(20):10023-10038

Ma ZS (2018b) Sketching the Human Microbiome Biogeography with DAR (Diversity-Area Relationship) Profiles. *Microbial Ecology*, 77(3): 821-838

Ma ZS, Li LW (2019) Semen microbiome biogeography: an analysis based on a Chinese population study. *Frontiers in Microbiology*, 9: 3333.

Plotkin JB, MD Potts, DW Yu, S Bunyavejchewin, R Condit, R Foster, S Hubbell et al. (2000) Predicting species diversity in tropical forests. *Proceedings of the National Academy of Sciences USA*, 97: 10850–10854.

Preston FW (1960) Time and space and the variation of species. *Ecology*. **41**: 611–627.

Renyi A (1961) On measures of entropy and information. In J. Neyman (Ed.), 4th Berkeley symposium on mathematical statistics and probability (pp. 547–561). Berkeley, CA.

Rosenzweig ML (1995) *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.

Segata N, Haake SK, Mannon P, et al (2012) Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol.* 13(6):R42.

Sizling AL, Kunin WE, Sizlingova E, Reif J, and Storch D (2011) Between geometry and biology: the problem if universality of the species-area relationship. *The American Naturalist*, 178(5): 602-611.

Storch D, Keil P & Jetz W (2012) Universal species–area and endemics–area relationships at continental scales. *Nature,* 488:78-83.

Tjørve E & Tjørve KMC (2008) The species–area relationship, self-similarity, and the true meaning of the z-value. *Ecology*, 89: 3528–3533.

Tjørve E (2009) Shapes and functions of species–area curves (II): a review of new models and parameterizations. *Journal of Biogeography*, 36: 1435–1445.

Triantis KA, Guilhaumon F and Whittaker RJ (2012) The island species–area relationship: biology and statistics. *Journal of Biogeography*, 39: 215–231

Ulrich W & Buszko J (2003) Self-similarity and the species–area relation of Polish butterflies. *Basic and Applied Ecology*, 4: 263–270.

Watson HC (1835) Remarks on geographic distribution of British plants. *London: Longman, Rees, Orme, Brown, Green and Longman*.

Wei Y, Zhu W, Gong J, et al. (2015) Fecal Microbiota Transplantation Improves the Quality of Life in Patients with Inflammatory Bowel Disease. Gastroenterol Res Pract., 517597. doi: 10.1155/2015/517597.

Weingarden AR, Vaughn BP (2017) Intestinal microbiota, fecal microbiota transplantation, and inflammatory bowel disease. *Gut Microbes*, 8(3):238-252. doi: 10.1080/19490976.

Figures

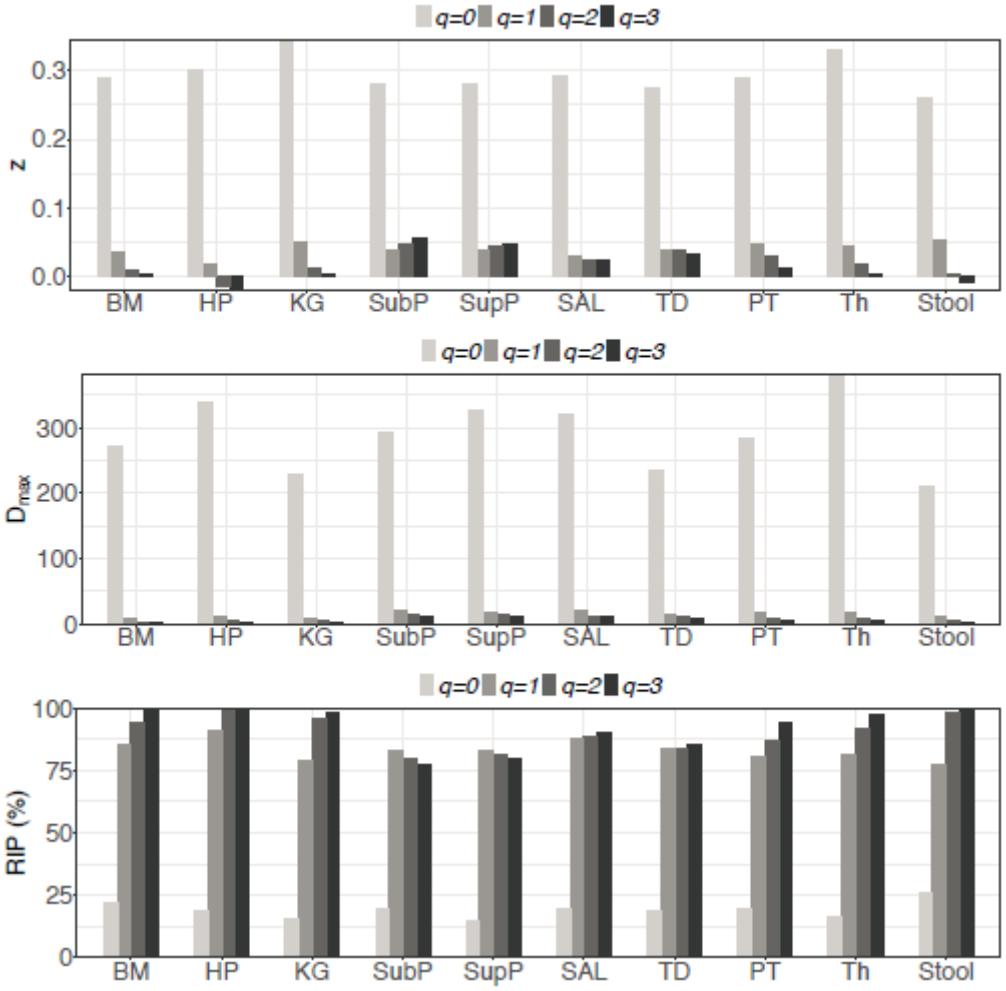

Figure 1

Graphs of three important profiles from inter-individual DAR models for each of 10 DT microbiome sites at genus taxon level, including DAR profiles (z-q patterns), MAD profiles (Dmax-q patterns), and RIP profiles (RIP-q patterns). Bar color indicates diversity order. The xaxis shows the DT sites: buccal mucosa (BM), keratinized gingiva (KG), hard palate (HP), throat (Th), palatine tonsils (PT), tongue dorsum (TD) and saliva (Sal), supraginval (SupP), subgingival plaques (SubP), and stool (Stool).

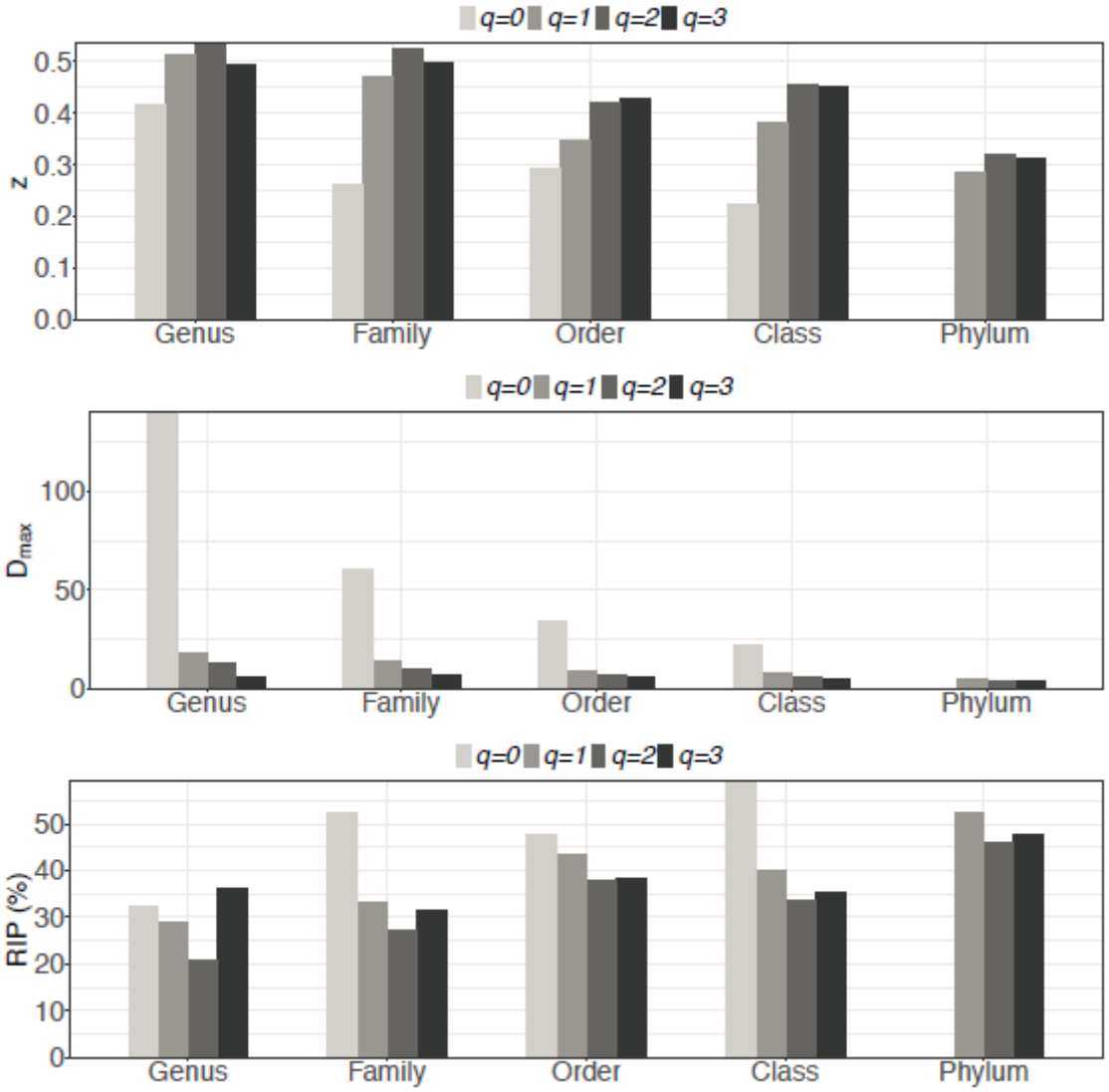

Figure 2

Graphs of three important profiles from intra-individual DAR models for each of 5 taxon levels, including DAR profiles (z-q patterns), MAD profiles (Dmax-q patterns), and RIP profiles (RIP-q patterns). Bar color indicates diversity order.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GutDAROSIV1.pdf