



Population-level diversity-disease relationship (p -DDR) in the human microbiome associated diseases

Wendy Li^{a,c}, Zhanshan (Sam) Ma^{a,b,c,*}

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

^b Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

^c Kunming College of Life Sciences, University of Chinese Academy of Sciences, China



ARTICLE INFO

Article history:

Received 9 November 2020

Received in revised form 9 April 2021

Accepted 13 April 2021

Available online 20 April 2021

Keywords:

Diversity-disease relationship (DDR)

Population-level diversity-disease relationship (p -DDR)

Diversity-area relationship (DAR)

Microbiome-associated disease (MAD)

Species-area relationship (SAR)

Potential microbiome diversity

Dark biodiversity

ABSTRACT

Diversity-disease relationship (DDR) is a *de facto* standard analysis in the studies of human microbiome associated diseases (MADs). For example, the species richness or Shannon entropy are routinely compared between the healthy and diseased groups. Nevertheless, the basic scale of the standard diversity analysis is individual subject rather than a cohort or population because the diversity is computed for individual samples, not for the group. Here we aim to expand the current DDR study from individual focus to population level, which can offer important insights for understanding the epidemiology of MADs. We analyzed the diversity-disease relationship at cohort scale based on a collection of 23 datasets covering the major human MADs. Methodologically, we harness the power of a recent extension to the classic species-area relationship (SAR), i.e., the diversity-area relationship (DAR), to achieve the expansion from individual DDR to inter-subject diversity scaling analysis. Specifically, we apply the DAR analysis to estimate and compare the potentially maximal accrual diversities of the healthy and diseases groups, as well as the inter-subject diversity scaling parameters and the individual-to-population diversity ratios. It was shown that, except for the potential diversity (D_{max}) at the cohort level in approximately 5.4% cases of MADs, DAR parameters displayed no significant differences between healthy and diseased treatments. That is, the DAR parameters are rather resilient against MADs, except for the potential diversity in some diseases. We compared our population-level DDR with the existing individual-level DDR patterns and proposed a hypothesis to interpret their differences.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Species diversity indexes such as species richness and Shannon entropy are routinely computed in virtually all studies of the human microbiome associated diseases (MADs). However, a recent *meta-analysis* [19] suggested that the diversity-disease relationship (DDR) is far less consistent than commonly perceived. In only approximately 1/3 of the analyzed cases, the DDR exhibited statistically significant disease effects, and in majority of the cases (approximately 2/3), there was not a consistent DDR relationship. Here, we further expand the current DDR study from individual focus [19] to cohort or population level (p -DDR).

The diversity scaling is a classic topic in biogeography that studies the spatial and temporal distribution of biodiversity [9].

* Corresponding author at: Computational Biology and Medical Ecology Lab, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, China.

E-mail address: ma@vandals.uidaho.edu (Zhanshan (Sam) Ma).

Medical geography combines biogeography with medicine, whose aim is to produce distribution patterns in the world biota by determining the interactions between multiple processes [21]. The macro-scale or biogeography perspective of disease can help us better describe, explain and predict the occurrence and development of diseases at community level. However, with the rise of molecular biology, medical geography has been gradually fading from our focus, even though it has a long and rich history [21]. Recently, Murray et al. [22], [23] reintroduced the theory and application of biogeography in the study of human infectious diseases, and referred it as pathogeography. Murray et al. [22] sketched the biogeographic map of 187 human infectious diseases across 225 countries, and found that there are distinct spatial patterns of human infectious diseases across the globe. They further reviewed the pathogeography and developed a framework for the study and management of human infectious diseases at macro-scale, heterogeneous-scale or biogeography level [21]. The DDR relationship explored in this study can be an important

supplement to Murray et al. [21] pathogeography framework, with examples from the field of human microbiome associated diseases (most of which are non-infectious diseases).

While the focus of the previous DDR study by Ma et al. [19] was to investigate whether two different individuals, one from the healthy (*H*) treatment (group) and another from the diseased (*D*) group (treatment) has significant difference in their microbiome diversities, in the present study, we are interested in whether or not two treatments, each as a whole, are significantly influenced by the MAD in terms of their cohort or population level diversity characteristics. For example, whether or not the healthy and diseased cohorts possess the same number of potential number of species (OTUs)? For another example, whether or not the intersubject scaling (changes) of diversity differs between the *H* & *D* cohorts? The expansion is significant because it extends the investigation from individual to cohort (population) level, and the scaling parameters may offer insights into the epidemiology of MADs—how diversity scaling is influenced by MADs in a population. Such scaling parameters are also of practical significance for devising public polices related to MADs.

Methodologically, we adopted the diversity-area relationship (DAR) [13,14,15,11], which is a recent extension to the classic SAR (species-area relationship) [27,3,28,5,12,4,32,31,6,29,30,33,36,7]. The SAR describes the relationship between the number of

species (*S*) and the number of individuals (or the so-termed “areas” *A*) accrued within a cohort or population (or the region), which can be described by a power law relationship. With the DAR analysis, one can estimate the potentially maximal accrual diversity in a cohort or population (D_{max}), the ratio of individual-to-population diversity (*RIP*), and also the parameter of diversity scaling (*z*) that can be considered as a measure of the intersubject heterogeneity in their microbiome diversities. The D_{max} offers estimates for the potential microbial diversity, also known as “dark” diversity in a cohort (or population), and rigorous statistical tests can be performed to compare two or more cohorts (*H* & *D* in the case of this study). For example, one may postulate that the number of opportunistic pathogens in the diseased treatment could rise, and the statistical test of D_{max} would be able to generate the answer.

Therefore, by adopting the DAR analysis, we are to deepen our understanding of the DDR in the human MADs by looking into the disease-associated changes at a larger scale of cohort (or population) beyond the *de facto* standard scale of individual subject in the current diversity analysis (as shown in Fig. 1). The objective of this study is then to investigate whether or not the *p*-DDR in MADs has similar patterns as in the case of individual level DDR previously investigated by Ma et al. [19]. We also propose a mechanistic hypothesis to interpret the observed *p*-DDR patterns.

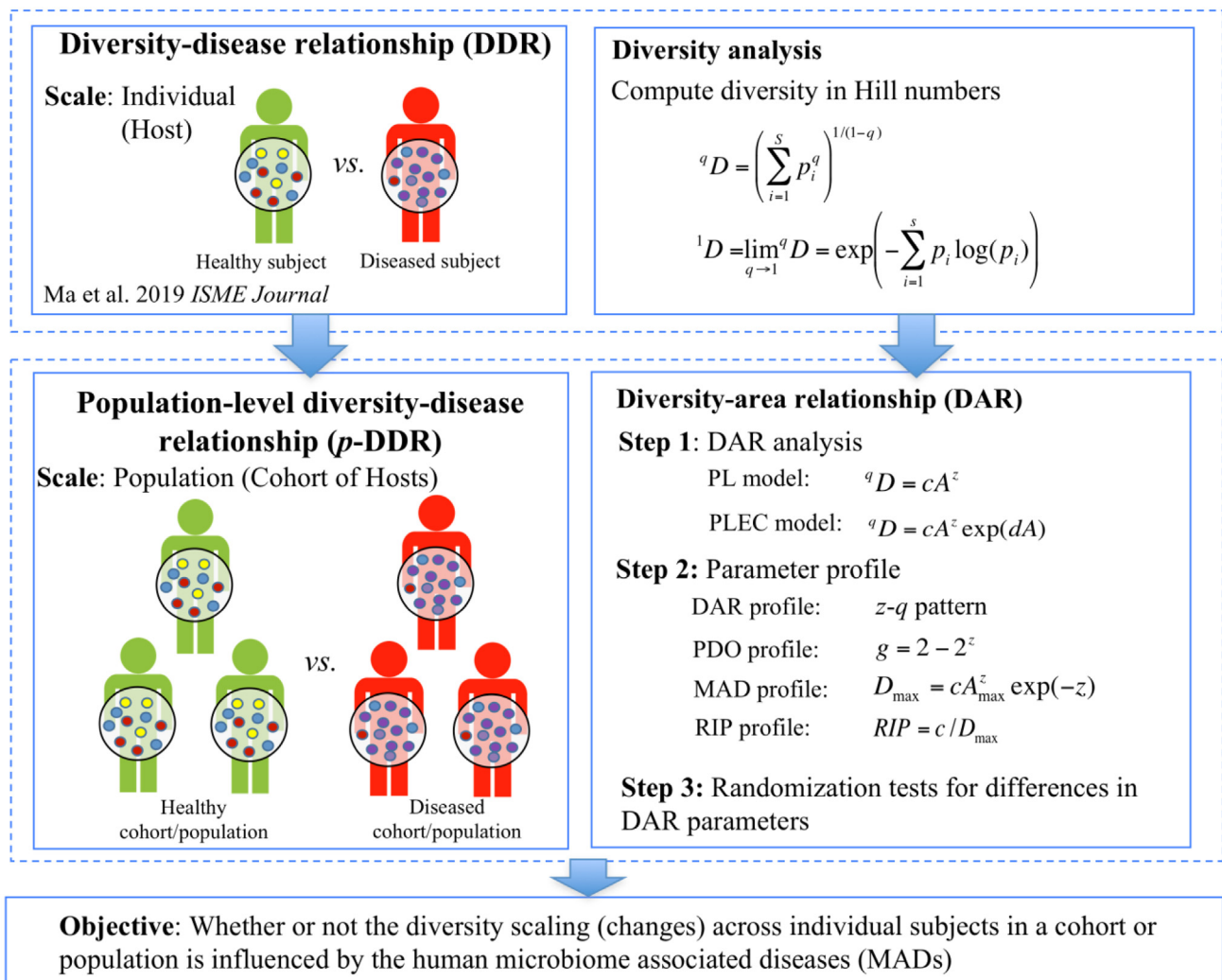


Fig. 1. A diagram showing the framework for investigating population-level diversity-disease relationship.

2. Material and Methods

2.1. The 23 datasets of the human microbiome associated diseases

The twenty-six 16s-rRNA datasets of the human microbiome associated diseases (MADs), which cover 7 human microbiome habitats (gut, oral, respiratory tract, skin, vaginal, semen and milk) and 16 diseases, were used in the diversity-disease relationship (DDR) study [19,15,16,17]. These 16 diseases included most of the high profile MADs such as obesity, IBD (inflammatory bowel disease), diabetes, autism, and schizophrenia. A brief introduction on those 23 datasets was presented in Table S1 of the online supplementary information (OSI).

2.2. The diversity-area relationship (DAR)

The DAR (diversity-area relationship) is an extension to the classic SAR (species-area relationship) and the history of the latter can be traced back to the 19th century [35], two decades earlier than Darwin's "Original of Species" was published. The SAR was considered as one of the most important laws in biogeography and conservation biology (ecology). For example, it plays a critical role in setting up the size of conservation region because the power law of SAR is rightly about the relationship between the number of species (S) in an area and the size of area (A), with a power function form: $S = cA^z$, where c and z are the parameters of the power function. Ma [13] extended the classic SAR to the general diversity-area relationship (DAR) by replacing the number of species (S) with the Hill numbers in the power law function. The Hill numbers are a series of diversity measures corresponding different diversity orders (q), weighted differently with species abundance frequency (distribution) in terms of q [8,1,2]. When $q = 0$, the Hill number defaults to the number of species or species richness (S). When $q = 1$, the Hill number (1D) is an exponential function of familiar Shannon entropy and can be considered as the number of typical or common species in the community. When $q = 2$, the Hill number (2D) is the reciprocal of Simpson index. Generally, qD is the diversity of a community consisting of $x = {}^qD$ equally abundant species. The Hill numbers are computed with the following formula:

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

where S is number of species, p_i is the relative abundance of species i , q is the order number of diversity.

In this study, we fitted two DAR models to the datasets of the human microbiome associated diseases (MADs). The first DAR model Ma (2018a, 2018b) selected and tested is the traditional power law (PL) model originally used for SAR modeling, i.e.,

$${}^qD = cA^z \quad (2)$$

where qD is diversity measured in the q -th order Hill numbers, A is area, and c & z are PL parameters. Based on the parameter z , we can estimate the pair-wise diversity overlap (PDO or g) of two bordering areas of the same size as:

$$g = 2 - 2^z \quad (3)$$

Another DAR model is the power law with exponential cutoff (PLEC), also originally used for modeling SAR [26,34,31], which has the following form:

$${}^qD = cA^z \exp(dA) \quad (4)$$

where d is a third parameter with taper-off effect. The item $\exp(dA)$ introduces the exponential decay to the original power law function

(eqn. (1)), which ultimately overwhelms the power law behavior when A becomes very large. Because both the human body and microbial species inhabited on or in human body are finite, the taper-off item in the PLEC should be justified. To fit the PL and PLEC models with Hill numbers, the following equations (5)(6) (log-linear transformation) will be used to estimate the model parameters of Eqns. (2)(4), respectively:

$$\ln(D) = \ln(c) + z \ln(A) \quad (5)$$

$$\ln(D) = \ln(c) + z \ln(A) + dA \quad (6)$$

In addition, a series of z -value in the DAR-PL model at different diversity orders ($q = 0, 1, 2, 3$) are termed as DAR profile [13].

2.3. Estimating the maximal accrual diversity profile with DAR-PLEC models

Ma [13] defined the concept of maximal accrual diversity of a cohort (or population) and derived its computational formula based on the PLEC model [eqn. (4) & (6)] as follows:

$$\text{Max}({}^qD) = {}^qD_{\max} = c \left(-\frac{z}{d} \right)^z \exp(-z) = cA_{\max}^z \exp(-z) \quad (7)$$

and the number of individuals (A_{\max}) required to reach the maximum can be estimated by

$$A_{\max} = -z/d \quad (8)$$

where all the parameters are the same as in eqns. (4) & (6).

The maximal accrual diversity profile ($D_{\max-q}$ pattern) was defined as a series of D_{\max} values at different diversity order (q) [13]. It is also a measure of "dark diversity" or potential diversity, which accounts for the species (diversity) locally absent but present in a habitat-specific regional species pool [23,24,10,27,20].

2.4. Estimating the RIP (the ratio of individual-to-population accrual diversity) profile

The RIP (Ratio of Individual diversity to Population accrual diversity) was defined as [18]:

$${}^qRIP = {}^q c / {}^q D_{\max} \quad (9)$$

where ${}^q c$ is the DAR-PL (eqn. (2)) parameter at diversity order of q , and ${}^q D_{\max}$ is the estimated maximal accrual diversity of the population (cohort) with eqn. (7) at diversity order q . The ${}^qRIP-q$ series (there is a RIP for each q) is known as the RIP profile. The RIP indicates the average level of an individual can represent a population (or cohort) from which the individual comes from. Since (D_{\max}) can be considered as a proxy of potential diversity, RIP is also related to the concept of potential diversity.

2.5. Testing the differences in the DAR parameters between H & D treatments

The permutation (randomization) test was utilized to test the differences in the parameters of DAR models between the H & D treatments. The null hypothesis (H_0) of permutation test is that the difference of a DAR parameter between the H and D treatments exceeds that between two randomly mixed groups (treatments). In other words, the difference in the values of a parameter is significantly influenced by random effect only and there is not a treatment effect. The steps of permutation test can be summarized as the following steps:

- (i) Computing the absolute values of the difference of DAR parameters between the H and D treatments, e.g., for parameter z , referred to as the true or observed difference (Δ_z).
- (ii) Pooling together all samples from the H and D treatments.
- (iii) Randomly reassigning the pooled samples into two groups generating the new H and D groups. The numbers of samples of new H and D groups remain the same with that of corresponding observed treatments in step (i). Fitting the DAR models based on the new H and D datasets. Getting the absolute values of the difference of their parameters between these two groups, and these differences are referred to as the expected or simulated differences (Δ_{sz}).
- (iv) Repeating step (iii) for 1000 times, we will get 1000 sets of expected differences.
- (v) For each parameter of DAR model, computing the p -value, which is defined as,

$$p = D_{\Delta_{sz} > \Delta_z} / 1000 \quad (10)$$

where $D_{\Delta_{sz} > \Delta_z}$ is the number of times when the expected (simulated) differences exceeds the observed difference. If p -value ≤ 0.05 , we can reject the null hypothesis and accept the alternative hypothesis, i.e., the treatment effect is significant or that the difference in parameter z is due to disease effects.

3. Results and discussion

In this study, we fitted two DAR models to the datasets of the human MADs, i.e. DAR-PL and DAR-PLEC models. The difference between DAR-PL and DAR-PLEC models are two-fold. First, the latter is an extension of the former by introducing a taper-off (or exponential cutoff) parameter, which sets a maximum for the accumulation of diversity. This extension should be more realistic because diversity on the earth planet is not unlimited. Second, the PLEC model allows one to estimate the potential (dark) diversity, which considers the contributions of species that are absent locally but present regionally (globally). The regional (global) species may act as potential sources for immigration to local communities.

Table S2 in the online supplementary information (OSI) lists the results of fitting DAR-PL and DAR-PLEC models to the 23 MAD datasets. Table S3 lists the results from the permutation tests for the differences in the DAR parameters between the healthy (H) and diseased (D) treatments. Figs. 2–5 illustrate the same information contained in these tables. In addition, Table 1 below summarizes the range of DAR parameters for the H & D treatments, respectively. From these results, we summarize the following findings:

(i) *DAR profile*: Diversity scaling parameter or rate (z) measures the inter-individual differences in species richness (Hill number for $q = 0$) or community diversity (Hill number for $q > 1$) of the microbiome for a population or cohort. A fast (rate) or larger z -value suggests large difference in diversity among individuals, or the higher heterogeneity among individuals. The DAR profile describes the relationship between scaling parameter (z) and diversity order (q). As shown in Fig. 2 and Table S2, the average scaling parameter (z) of the H treatment across diversity order $q = 0$ –3 is $z = (0.558, 0.387, 0.293, 0.248)$, and the average z value of the D treatment across diversity order $q = 0$ –3 is $z = (0.519, 0.344, 0.275, 0.236)$. In either treatment, the scaling parameter (z) decreases with the diversity order (q), which is determined by the nature of Hill numbers—higher order diversity is weighted less by common species. Therefore, at higher diversity order, the scaling (scale of change) parameter becomes smaller.

(ii) *PDO (pair-wise diversity overlap) profile*: PDO and diversity scaling are like two sides of a coin, in which later describes the inter-individual difference, and the former describes the inter-individual similarity. As shown in Fig. 3 and Table S2, the average PDO parameter (g) of the H treatment across diversity order $q = 0$ –3 is $g = (0.511, 0.668, 0.745, 0.784)$, and the average g value of the D treatment across diversity order $q = 0$ –3 is $g = (0.555, 0.712, 0.766, 0.799)$. The PDO profiles of both H and D treatments monotonically increased with q . This is again determined by the nature of Hill numbers, as explained previously, because at high diversity orders, the scaling (scale of change) parameter becomes smaller and consequently, the overlap (g) parameter rises.

(iii) *Maximal accrual diversity profile*: As mentioned in the section of Methods, maximal accrual diversity or parameter D_{max} equals to the potential microbial biodiversity for a human population or cohort. For example, when $q = 0$, D_{max} estimates the maximum number of species that a population can have. When $q > 1$, D_{max} estimates the maximum number of species with higher level of commonness of a cohort. In the case of this study, it measures the influences of other individuals in a population on the microbial diversity of a specific individual (an “average Joe”). As shown in Fig. 4 and Table S2, the average D_{max} of the H treatment across diversity order $q = 0$ –3 is $D_{max} = (2495.5, 625.7, 62.2, 39.0)$, and the average D_{max} of the D treatment across diversity order $q = 0$ –3 is $D_{max} = (2418.9, 256.8, 62.7, 40.3)$. The maximal accrual diversity profiles of both H and D treatments decreased monotonically decreasing with q . This, of course, is determined by the nature of Hill numbers.

(iv) *RIP (ratio of individual diversity to population accrual diversity) profile*: RIP can be used to estimate the percentage of population diversity represented by an average individual. As shown in Fig. 5 and Table S2, the average RIP of the H treatment across diversity order $q = 0$ –3 is $RIP = (20.2, 31.3, 51.3, 60.4)$, and the average RIP of the D treatment across diversity order $q = 0$ –3 is $RIP = (21.3, 37.0, 44.5, 53.2)$. The RIP profiles of both H and D treatments monotonically increased with q .

(v) As shown in Table S3, only 8 out of 148 comparisons (5.4%) in the D_{max} parameter showed significant differences between the H & D treatments. However, virtually no significant differences were detected for any of the other parameters. There was only one comparison of PDO parameter (g) showed significant difference between the H & D in the case of BV at diversity order $q = 0$.

In summary, the previous results indicate that all diversity-area relationship (DAR) parameters, except for D_{max} parameter, are not significantly influenced by the microbiome-associated diseases. Even in the exceptional case of D_{max} , only in 5.4% cases, D_{max} showed significant differences between the H & D treatments. That is, diseases may significantly influence the cohort and population level potential diversity.

Given that the individual-level diversity-disease relationship (DDR) was influenced significantly by diseases in only approximately 1/3 of the cases, the lack of disease effects on most DAR parameters should not be surprising. We postulate the following interpretations for the findings revealed previously. First, the previous individual level DDR already revealed that in the majority of cases (2/3), diseases did not have significant effects on individual level diversity [19]. Obviously, it is unlikely that in the same majority cases, diseases could have significant effects on the cohort or population level DDR. Second, as to the difference between the previous individual-level DDR and this study in the minority of cases, i.e., with effects in 1/3 of the cases vs. virtually 0 case in all DAR parameters except for D_{max} , we postulate that, at the cohort or population level, the difference in diversity may be canceled



Fig. 2. Graphs of the DAR-PL scaling parameter (z) at different diversity order $q = 0, 1, 2, \& 3$ (DAR profile), for the healthy (H) and diseased (D) treatments of the 23 case studies. Light green and dark green indicate healthy cohort, orange and brown indicate diseased cohort. Note that most datasets only have one healthy or diseased state, but some may have two, then all color bars were necessary. See Table S2 (in Supporting Information) for the numeric values of the parameter z . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

each other at the cohort level, due to the inter-subject heterogeneity (differences) in their individual (base) level diversities. In other words, some individuals may have inherently high diversity than others, and others may have inherently lower diversity than the

other. A net effect is then the reduction of the differences at the cohort or population level.

As to why the D_{max} showed 5.4% disease effects, which is still lower than the 1/3 of the cases in previous individual-level DDR

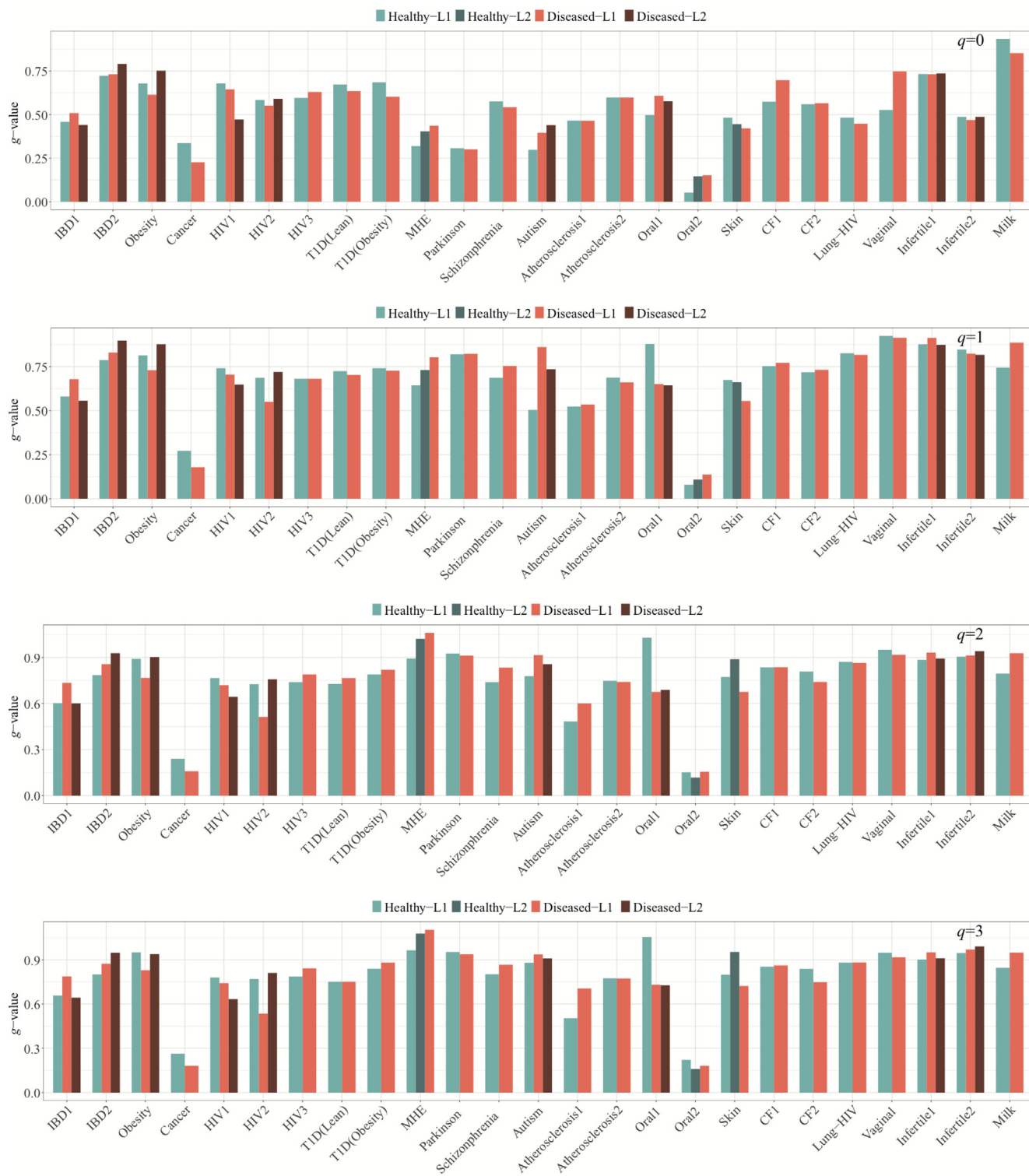


Fig. 3. Graphs of the PDO (pair-wise diversity overlap) profile (g - q series): g corresponding to different diversity order $q = 0, 1, 2, \& 3$ for the healthy (H) and diseased (D) treatments of the 23 case studies. Light green and dark green indicate healthy cohort, orange and brown indicate diseased cohort. Note that most datasets only have one healthy or diseased state, but some may have two, then all color bars were necessary. See Table S2 (in Supporting Information) for the numeric values of the parameter g . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

pattern, but significantly higher than the disease effects on the other DAR parameters (which are zero except for one case), the answer can be found by examining the differences among the DAR parameters. The PDO parameter (g) measures the

pair-wise overlap (similarity) between two individuals. Similarly, the DAR scaling parameter (z) is also determined by two points (eqn. (4) is a straight line). In other words, both z and g could be determined by two individuals or by the pair-wise heterogeneity.



Fig. 4. Graphs of the maximal accrual diversity profile (D_{max-q} series): D_{max} corresponding to different diversity order $q = 0, 1, 2, & 3$ for the healthy (H) and diseased (D) treatments of the 23 case studies. Light green and dark green indicate healthy cohort, orange and brown indicate diseased cohort. Note that most datasets only have one healthy or diseased state, but some may have two, then all color bars were necessary. Comparisons with significant differences in D_{max} between H & D treatments were marked with asterisks. See Table S2 (in Supporting Information) for the numeric values of the parameter D_{max} . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In contrast, D_{max} is a ‘global’ parameter given that it is the ‘total’ diversity owned by the whole cohort or population, rather than determined by a pair of two individuals. Obviously, it should be

much more difficult to cancel the differences globally that pairwise. This explains why D_{max} preserved certainly level of differences in the disease effects passed up from individual-level diver-



Fig. 5. Graphs of the RIP (ratio of individual diversity to population accrual diversity) profile ($RIP-q$ series): RIP corresponding to different diversity order $q = 0, 1, 2, \& 3$ for the healthy (H) and diseased (D) treatments of the 23 case studies. Light green and dark green indicate healthy cohort, orange and brown indicate diseased cohort. Note that most datasets only have one healthy or diseased state, but some may have two, then all color bars were necessary. See Table S2 (in Supporting Information) for the numeric values of the parameter RIP . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
The ranges of DAR (diversity-area relationship) parameters for the healthy and diseased treatments, respectively, averaged from 23 MAD case studies exhibited in Tables S1–S2.

| Order | Treatment | Statistics | PL (Power Law) | | | PLEC (Power Law with Exponential Cutoff) | | | | | | |
|---------|-----------|------------|----------------|--------|-------|--|--------|--------|------------------|------------------|---------|-------|
| | | | z | ln(c) | g | z | d | ln(c) | A _{max} | D _{max} | RIP (%) | |
| q = 0 | Healthy | Mean | 0.558 | 5.391 | 0.511 | 0.724 | −0.020 | 5.251 | 51.859 | 2495.5 | 20.2 | |
| | | Min | 0.092 | 2.219 | 0.053 | 0.170 | −0.087 | 1.948 | 0.000 | 109.2 | 1.4 | |
| | | Max | 0.960 | 7.358 | 0.934 | 1.013 | 0.000 | 7.169 | 195.797 | 10663.0 | 79.9 | |
| | Disease | Std. Err. | 0.034 | 0.210 | 0.035 | 0.036 | 0.003 | 0.211 | 8.049 | 442.516 | 2.849 | |
| | | Mean | 0.519 | 5.567 | 0.555 | 0.667 | −0.018 | 5.437 | 63.317 | 2418.9 | 21.3 | |
| | | Min | 0.196 | 3.367 | 0.152 | 0.254 | −0.054 | 3.219 | 0.000 | 124.9 | 4.1 | |
| | q = 1 | Healthy | Max | 0.885 | 7.577 | 0.853 | 1.021 | 0.001 | 7.334 | 362.256 | 8792.3 | 66.3 |
| | | | Std. Err. | 0.027 | 0.170 | 0.027 | 0.031 | 0.002 | 0.166 | 13.411 | 365.281 | 1.969 |
| | | | Mean | 0.387 | 3.475 | 0.668 | 0.515 | −0.015 | 3.358 | 99.682 | 625.7 | 31.3 |
| Disease | | Min | 0.102 | 0.855 | 0.080 | 0.135 | −0.048 | 0.719 | 0.000 | 3.6 | 1.2 | |
| | | Max | 0.938 | 4.990 | 0.925 | 0.971 | 0.005 | 4.906 | 1149.590 | 6398.3 | 65.3 | |
| | | Std. Err. | 0.039 | 0.173 | 0.039 | 0.043 | 0.003 | 0.170 | 41.003 | 298.537 | 2.891 | |
| q = 2 | | Healthy | Mean | 0.344 | 3.570 | 0.712 | 0.470 | −0.013 | 3.426 | 53.032 | 256.8 | 37.0 |
| | | | Min | 0.116 | 1.542 | 0.138 | 0.086 | −0.034 | 1.410 | 6.517 | 10.5 | 2.8 |
| | | | Max | 0.896 | 4.930 | 0.914 | 0.993 | 0.018 | 4.874 | 192.343 | 3466.4 | 77.0 |
| | Disease | Std. Err. | 0.031 | 0.126 | 0.030 | 0.036 | 0.002 | 0.127 | 7.739 | 102.446 | 2.749 | |
| | | Mean | 0.293 | 2.636 | 0.745 | 0.398 | −0.012 | 2.538 | 534.429 | 62.2 | 51.3 | |
| | | Min | −0.085 | 0.662 | 0.118 | −0.172 | −0.038 | 0.538 | 0.000 | 2.6 | 3.3 | |
| | q = 3 | Healthy | Max | 0.908 | 3.948 | 1.029 | 0.938 | 0.011 | 3.888 | 14.091 | 524.0 | 129.4 |
| | | | Std. Err. | 0.047 | 0.144 | 0.044 | 0.050 | 0.003 | 0.142 | 502.120 | 20.842 | 5.220 |
| | | | Mean | 0.275 | 2.673 | 0.766 | 0.371 | −0.009 | 2.561 | 82.883 | 62.7 | 44.5 |
| Disease | | Min | −0.127 | 1.013 | 0.156 | −0.001 | −0.033 | 0.902 | 0.000 | 4.7 | 2.1 | |
| | | Max | 0.881 | 3.953 | 1.061 | 0.956 | 0.009 | 3.935 | 428.557 | 526.1 | 89.6 | |
| | | Std. Err. | 0.036 | 0.106 | 0.034 | 0.038 | 0.002 | 0.107 | 16.934 | 16.695 | 3.569 | |
| q = 3 | | Healthy | Mean | 0.248 | 2.319 | 0.784 | 0.348 | −0.012 | 2.226 | 32.493 | 39.0 | 60.4 |
| | | | Min | −0.163 | 0.579 | 0.160 | −0.250 | −0.041 | 0.470 | 0.000 | 2.4 | 3.4 |
| | | | Max | 0.874 | 3.467 | 1.077 | 0.907 | 0.014 | 3.432 | 132.299 | 348.7 | 164.0 |
| | Disease | Std. Err. | 0.048 | 0.129 | 0.044 | 0.048 | 0.003 | 0.128 | 5.470 | 13.694 | 6.230 | |
| | | Mean | 0.236 | 2.346 | 0.799 | 0.323 | −0.010 | 2.243 | 54.200 | 40.3 | 53.2 | |
| | | Min | −0.188 | 0.874 | 0.182 | −0.053 | −0.041 | 0.765 | 0.000 | 3.6 | 1.8 | |
| | Disease | Max | 0.859 | 3.582 | 1.103 | 0.912 | 0.018 | 3.553 | 234.313 | 431.9 | 101.0 | |
| | | Std. Err. | 0.037 | 0.096 | 0.034 | 0.038 | 0.002 | 0.096 | 11.078 | 13.340 | 4.052 | |

sity difference. As to the RIP parameter, the ratio nature is again easy to cause the cancelation of the differences in disease effects, and hence, displayed similar behavior as z and g.

CRediT authorship contribution statement

Wendy Li: Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Zhanshan (Sam) Ma:** Conceptualization, Methodology, Investigation, Writing - original draft, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We appreciate Dr. LW Li of the Chinese Academy of Sciences for his computational support to this study.

Data availability.

All datasets analyzed in this study are available in public domain and see Table S1 for the detailed access information for each of the 23 datasets.

Funding

This work received funding from the National Natural Science Foundation of China (NSFC) Grant (No. 31970116) on Medical Ecology of Human Microbiome.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2021.04.032>.

References

- Chao A, Chiu CH, Hsieh TC. Proposing a resolution to debates on diversity partitioning. *Ecology* 2012;93(9):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., Evolut. Systemat.* 2014;45:297–324.
- Connor EF, McCoy ED. The statistics and biology of the species-area relationship. *Am. Nat.* 1979;113:791–833.
- Drakare S, Lennon JJ, Hillebrand H. The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecol. Lett.* 2006;9:215–27.
- Harte J, Kinzig A, Green J. Self-similarity in the distribution and abundance of species. *Science* 1999;284:334–6.
- He F, Hubbell SP. Species-area relationships always overestimate extinction rates from habitat loss. *Nature* 2011;473:368–71.
- Helmus MR, Mahler DL, Losos JB. Island biogeography of the Anthropocene. *Nature* 2014;513:543–7.
- Jost L. Partitioning diversity into independent alpha and beta components. *Ecology* 2007;88:2427–39.
- Ladau J, Eloie-Fadrosch EA. Spatial, Temporal, and Phylogenetic Scales of Microbial Ecology. *Trends Microbiol.* 2019;27(8):662–9.
- Li K, Bihan M, Yooseph S, Methe BA. Analyses of the microbial diversity across the human microbiome. *PLoS ONE* 2012;7:e32118.
- Li W, Ma ZS. Diversity scaling of the human vaginal microbial communities. *Zool. Res.* 2019. <https://doi.org/10.24272/j.issn.2095-8137.2019.068>.
- Lomolino MV. Ecology's most general, yet protean pattern: the species-area relationship. *J. Biogeogr* 2000;27:17–26.
- Ma ZS. DAR (diversity-area relationship): Extending classic SAR (species-area relationship) for biodiversity and biogeography analyses. *Ecol. Evol.* 2018;8(20):10023–38.
- Ma ZS. Sketching the Human Microbiome Biogeography with DAR (Diversity-Area Relationship) Profiles. *Microb. Ecol.* 2018;77(3):821–38.
- Ma ZS. Critical network structures and medical ecology mechanisms underlying human microbiome-associated diseases. *iScience* 2020. <https://doi.org/10.1016/j.isci.2020.101195>.

- [16] Ma ZS. Heterogeneity-Disease relationship in the human microbiome associated diseases. *FEMS Microbiol. Ecol.* 2020;fiae093. <https://doi.org/10.1093/femsec/fiae093>.
- [17] Ma ZS. Testing the Anna Karenina Principle in human microbiome-associated diseases. *iScience* 2020;23(4):101007.
- [18] Ma ZS, Li LW. Semen microbiome biogeography: an analysis based on a Chinese population study. *Front. Microbiol.* 2019;9:3333.
- [19] Ma ZS, Li LW, Gotelli NJ. Diversity-disease relationships and shared species analyses for human microbiome-associated diseases. *ISME J.* 2019. <https://doi.org/10.1038/s41396-019-0395-y>.
- [20] Ma ZS. A new DTAR (diversity-time-area relationship) model demonstrated with the indoor microbiome. *J. Biogeogr.* 2019;00:1–18.
- [21] Murray KA, Olivero J, Roche B, et al. Pathogeography: Leveraging the biogeography of human infectious diseases for global health management. *Ecography* 2018;41:1411–27.
- [22] Murray KA, Preston N, Allen T, et al. Global biogeography of human infectious diseases. *Proc. Natl Acad. Sci. USA* 2015;112:12746–51.
- [23] Partel M, Szava-Kovats R, Zobel M. Dark diversity: shedding light on absent species. *Tren. Ecol. Evolut.* 2011;26:124–8.
- [24] Partel M, Szava-Kovats R, Zobel M. Discerning the niche of dark diversity. *Trends Ecol. Evol.* 2011;26(6):265–6.
- [25] Plotkin JB, Potts MD, Yu DW, et al. Predicting species diversity in tropical forests. *Proceed. Natl. Acade. Sci. USA* 2000;97:10850–4.
- [26] Preston FW. Time and space and the variation of species. *Ecology* 1960;41:611–27.
- [27] Real R, Barbosa AM, Bull JW. Species Distributions, Quantum Theory, and the Enhancement of Biodiversity Measures. *Syst. Biol.* 2017;66(3):453–62.
- [28] Rosenzweig ML. *Species Diversity in Space and Time.* Cambridge: Cambridge University Press; 1995.
- [29] Sizing AL, Kunin WE, Sizingová E., et al. Between geometry and biology: the problem of universality of the species-area relationship. *Am. Nat.* 2011;178:602–11.
- [30] Storch D, Keil P, Jetz W. Universal species-area and endemics-area relationships at continental scales. *Nature* 2012;488:78–83.
- [31] Tjørve E. Shapes and functions of species-area curves (II): a review of new models and parameterizations. *J. Biogeogr.* 2009;36:1435–45.
- [32] Tjørve E, Tjørve KM. The species-area relationship, self-similarity, and the true meaning of the z-value. *Ecology* 2008;89:3528–33.
- [33] Triantis KA, Guilhaumon F, Whittaker RJ. The island species-area relationship: biology and statistics. *J. Biogeogr* 2012;39:215–31.
- [34] Ulrich W, Buszko J. Self-similarity and the species-area relation of Polish butterflies. *Basic Appl Ecol* 2003;4:263–70.
- [35] Watson HC. *Remarks on geographic distribution of British plants.* London: Longman, Rees, Orme, Brown, Green and Longman; 1835.
- [36] Whittaker RJ, Triantis KA. The species-area relationship: an exploration of that 'most general, yet protean pattern'. *J. Biogeogr* 2012;39:623–6.