Further Quantifying the Niche-Neutral Continuum of Human Digestive Tract Microbiomes with Near Neutral Model and Stochasticity Analysis

Evolutionary Bioinformatics
Volume 18: 1–12

DOI: 10.1177/11769343221128540
SAGE

Hongju (Daisy) Chen[1,2] and Zhanshan (Sam) Ma[2,3]

[1]Department of Mathematics and Statistics, Honghe University, Yunnan, China. [2]Computational Biology and Medical Ecology Lab, State Key Lab of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Science, Kunming, Yunnan, China. [3]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, China.

ABSTRACT: It is postulated that the human digestive tract (DT) from mouth to intestine is differentiated into diverse niches. For example, Segata et al. discovered that the microbiomes of diverse habitats along the DT could be distinguished as 4 types (niches) including (i) stool; (ii) subgingival plaques (SubP) and supra-gingival plaques (SupP); (iii) tongue dorsum (TD), throat (TH), palatine tonsils (PT), and saliva (Sal); and (iv) hard palate (HP) and buccal mucosa (BM), and keratinized gingiva (KG). These niches are different not only in composition, but also in metabolic potentials. In a previous study, we applied Harris et al's multi-site neutral and Tang and Zhou's niche-neutral hybrid models to characterize the DT niches discovered by Segata et al. Here, we complement the previous study by applying Sloan's near-neural model and Ning et al's stochasticity analysis framework to quantify the niche-neutral continuum of the DT microbiome distribution to shed light on the possible ecological/evolutionary mechanism that shapes the continuum. Overall but excluding the stool site, the proportion of neutral OTUs (46%) is slightly higher than that of the positive selection (38%), but significantly higher than negative selection (15%). The gut (stool) exhibited 3 to 12 times lower neutrality than other DT sites. The analysis also cross-verified our previous hypothesis that the KG (*keratinized gingiva*) is of distinct assembly dynamics in the DT microbiome, should be treated as a fifth niche. Our findings offer new insight on the long-standing debate concerning whether a minimum of 2-mm of KG width is necessary for marginal periodontal health.

KEYWORDS: Unified neutral theory of biodiversity (UNTB), digestive-tract (DT) niche differentiation, near-neutral model, normalized stochasticity ratio (NSR), keratinized gingiva (KG) niche, positively selected species, neutral species, negatively selected species

RECEIVED: February 9, 2022. **ACCEPTED:** August 28, 2022.

TYPE: Metagenomics and the Analysis of Microbiomes – Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study received funding from the National Natural Science Foundation (Grant #72274192; #31970116; # 12161033) and the State Key Laboratory of Genetic Resources and Evolution (GREKF19-07); Yunnan Province Local University (Part) Basic Research for Youths (No. 202001BA070001-100; No. 202101BA070001-018); The Yunnan provincial department of education scientific research fund project (No. 2022 J0896).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Zhanshan (Sam) Ma, Computational Biology and Medical Ecology Lab, State Key Lab of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Science, Kunming, Yunnan 650204, China.
Email: ma@vandals.uidaho.edu

Introduction

How ecological communities such as human microbiota (which can be roughly considered as the regional existence of human microbiomes) are assembled and how their diversities are maintained (without being monopolized by individuals of a single species) has been wondered, to the minimum, ever since the time of Darwin's "Origin of Species," and is still not fully settled until today. A central principle of Darwin's evolution by natural selection is that species compete for living in nature and it is intuitively plausible that the world could be dominated or even monopolized by a single species such as *Homo sapiens*. Nevertheless, the reality diametrically contradicts that obviously wrong imagination, that is, the nature is astonishingly diverse and supports numerous species of living organisms, and we humans by no means can dominate the natural world. The COVID-19 is a vivid example to demonstrate our competitive disadvantage. Our digestive tract (DT) is another example, which is believed to harbor the greatest microbiome diversity of our bodies (eg,[1-3]).

In the 1920s, naturalists and ecologists conceived that it may be the case that is the diversity (strictly speaking heterogeneity) of nature that makes it possible to support the astonishingly diverse species, that is, each species may live and prosper in its own species-specific niche. Of course, the above statement is an overly simplified or idealized view on the niche theory, which has been studied and advanced extensively during the last century or so (eg,[3,4]). In reality, niche is a rather vague concept, and it is hardly possible to define niches for many species, not to mention measure them quantitatively. For example, Deines et al[4] demonstrated that even with a single metaorganism *Hydra vulgaris* (strain AEP), discrepancies exist between the fundamental niches and realized niches of bacteria species that colonize *Hydra*. This suggests that there are other natural forces beyond niche selection that are in effects in determining the structure of communities.

Indeed, in the 1990s, Hubbell[5] had already formulated a systematic, alternative theory to the niche (selection) theory, well known as the "Unified neutral theory of biodiversity and biogeography" (UNTB). Hubbell's UNTB was inspired by Kimura's neutral theory of molecular evolution in population genetics, which maintains that most evolutionary changes such as gene mutations are neutral in the sense that they are caused by stochastic genetic drifts without phenotypic consequences. Similarly, the UNTB maintains that species in communities

are neutral in the sense that their differences are irrelevant to their successes; essentially niche differentiations do not influences species abundance, and the abundance of each species follows a random walk.[6,7] More specifically, neutrality means that measured on a per capita basis, species are equivalent in their demographic characteristics including birth, death, dispersal, and speciation rates. Although the neutrality assumption is frequently violated in reality, the UNTB offered an important starting point for understanding the effects of ecological *drift* (demographic stochasticity) among equivalent species with identical vital rates. More importantly, it can be formulated as a precise statistical model for testing, and frequently, rejecting, the null hypothesis that community structures are determined by stochasticity alone.

Like many fields of science, the truth is most likely in the middle of both diametrically opposite theories. During the last couple of decades, there have been several "hybrid models" that tried to reconcile both niche and neutral theories (eg,[8-17]). These hybrid models recognize the roles of both deterministic niche selection and stochastic neutral drifts, and often try to quantify their relative importance in structuring ecological communities. The so-termed niche-neutral continuum is such a concept that is often used to represent the spectrum between 2 extreme ends of total deterministic niche *vs*. complete stochastic drifts. However, to quantify the niche-neutral continuum is challenging. Two approaches, Sloan et al[18,19] near neutral model and Ning et al[20] stochasticity analysis framework appear to be particularly useful for dealing with the challenge. We will apply both the approaches in this study for quantifying the niche-neutral continuum of the human DT microbiomes.

According to Segata et al,[2] the DT is differentiated into 4 diverse habitat types or niches. Their study is important because it for the first time proposed the concept of niche differentiation of the DT microbiomes, although they failed to present quantified characterizations of the 4 niches or habitat types. In a previous study, Chen and Ma[21] provided a preliminary quantification of the niche types originally proposed by Segata et al[2] by applying Harris et al[22] multi-site neutral (MSN) model and Tang and Zhou[14] niche-neutral hybrid model. Both MSN and NNH models can be considered as further advances to Hubbell's UNTB, and both are assumed to measure the levels of stochastic neutrality across multiple sites of the DT microbiomes.[21]

The primary objective of this study is to further quantify the niche-neutral continuum of the DT microbiome beyond what is obtained from the MSN/NNH modeling.[21] Specifically, we apply Ning et al[20] framework for community stochasticity analysis, which offers a metric termed normalized stochasticity ratio (NSR) with a value between 0% and 100% or between 0 and 1. If the community assembly is extremely deterministic without any stochasticity, then NSR should be 0%; otherwise NSR should be 100%. Applied to niche-neutral continuum, the NSR approaches to 0 at the niche end, and approaches to 100% at the neutral end. While Ning et al[20] NSR metric was computed at the community (more accurately, meta-community scale), we applied Sloan et al[18,19] near-neutral model to perform an OTU (operational taxonomic unit) or species-level neutrality analysis, which can classify microbial species as neutral, positively selected and negatively selected species, similar to the neutral, positively and negatively selected genes in population genetics. Sloan et al's[18,19] model is termed near-neutral model because it slightly relaxed the neutrality assumption in Hubbell's[5] classic UNTB, but preserved the essential elements of the UNTB.

To fulfill the previously defined objective, in the remainder of this article, we present our integrated analyses of the human digestive tract (DT) microbiome with Sloan et al[18,19] near-neural model and Ning et al[20] stochasticity framework by reanalyzing a large HMP (human microbiome project) dataset previous published by Segata et al.[2] The integrated analyses produce quantitative characterization of the niche-neutral continuum of the DT microbiome, that is, the level of stochasticity versus the level of deterministic selection (in the form of niche differentiations).

Material and Methods

Multi-site microbiome datasets along the human digestive tract (DT)

Figure 1 illustrated the 10 sampling sites of the human DT (digestive tract) microbiomes,[2] as well as the mathematical models used to perform OTU-level neutrality/selection analysis[18,19] (near-neutral model) and community-level stochasticity analysis[20] (stochasticity framework). The diagram (Figure 1) also showed the roadmap to achieve our objective—characterizing the niche-neutral continuum of human DT microbiomes. In other words, we aim to gage the parameters that can be used to quantify the relative importance of stochastic neutral drifts versus deterministic niche differentiations (selection) in shaping the community structures (diversity patterns) of human microbiomes along the DT.

Segata et al[2] collected a total of 2105 microbiome samples from 10 sites of over 200 individual subjects long their DTs. They sequenced those samples with 16S-rRNA amplicon sequencing and computed the OTU (operational taxonomic unit) tables of the DT microbiomes. The 10 DT sites they sampled include and were grouped as 4 niche types, as follows: (i) buccal mucosa (BM), keratinized gingiva (KG), hard palate (HP); (ii) saliva (Sal), tongue dorsum (TD), palatine tonsils (PT), throat (TH); (iii) sub-gingival plaques (SubP) and supra-gingival plaques (SupP); and (iv) stool. These 4 DT niches were found to possess not only distinctive species community compositions, but also different metabolic potentials functionally. It should be noted that the *niche* concept we use in this article refers to niche for microbiota, rather than niche for specific species.[23] Similarly, the concept of selection we use in this article is also slightly different from its traditional meaning, natural selection for species fitness. Instead, we use the term

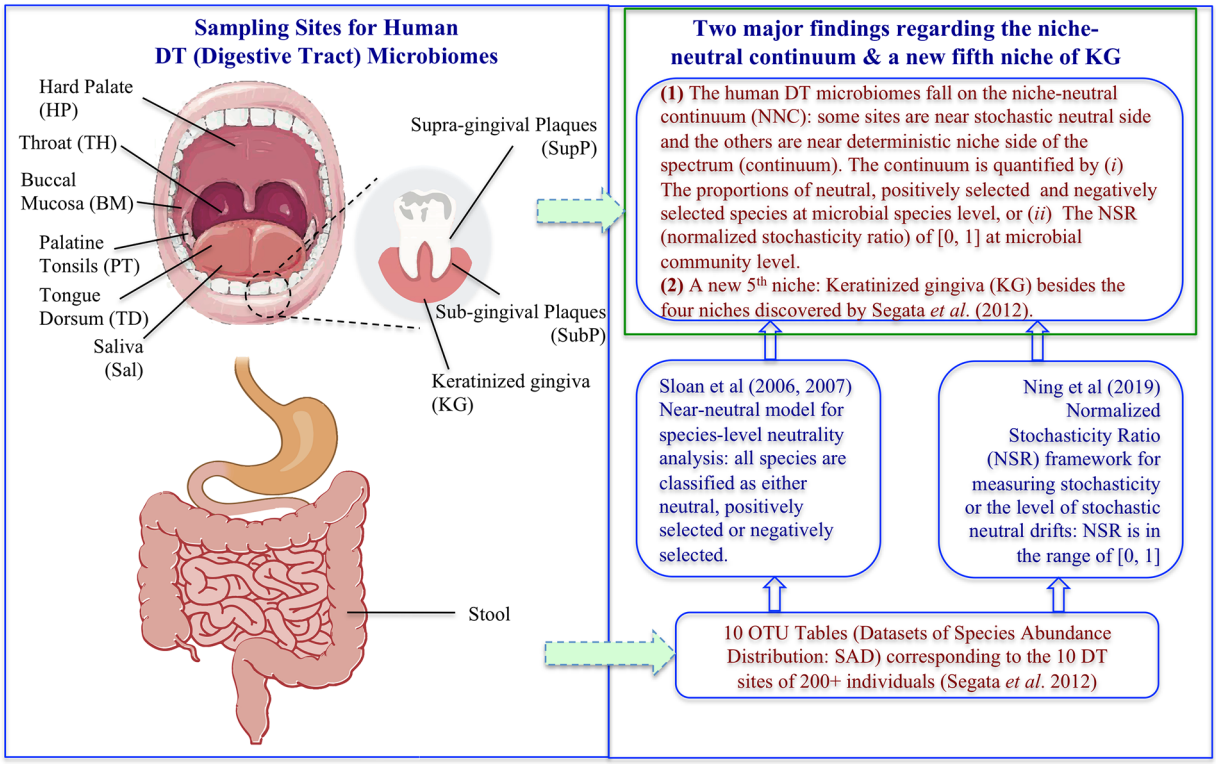

Figure 1. A diagram for illustrating the sampling sites of the human DT microbiome and analysis design.

selection to refer to niche selection or selection of microbiota niche. Furthermore, selection, in particular, in the context of selection in gut environment, can be multidimensional, where a high diverse of molecules originated either by the host or the microbiota influences positive or negative selection, termed complex eco-active intestinal chemosphere.[24]

Sloan et al[18,19] near-neutral model for the OTU-level neutrality analysis

Sloan et al[18,19] derived an alternative neutral model based on the principle of Hubbell's neutral theory. Sloan's neutral model was designed to address the challenges in inferring the species abundance distribution of an entire microbial community from disproportionally small magnitude of metagenomic samples (eg, a few grams of fecal sample vs the whole length intestine). The theoretic occurrence frequency of each species predicted from Sloan model is compared with the observed species abundances, and all species are classified into 3 categories: neutral, below-neutral, and above-neutral species, whose abundances are consistent (falling within the 95% confidence interval), below or above those predicted by the Sloan model. Unlike Hubbell[5] neutral model, there is not a community level test statistic for determining neutrality. We argue that the value of Sloan's model lies in its capacity to distinguish species as neutral, "positively selected" (above neutral), or "negatively selected" (below neutral). Another useful metric for evaluating the goodness-of-fitting of Sloan's model is the generalized R^2. For detailed information on Sloan model, readers are referred to Sloan et al[18,19] and Burns et al.[25] In this study, we use Sloan et al near-neutral model to gain OTU-level insights on the niche-neutral continuum of the DT microbiomes, while Ning et al[20] stochasticity analysis is used to gain community-level insights. Burns et al[25] software package that implemented Sloan et al[18,19] model is used in this study to perform OTU-level neutrality analysis. Sloan's near-neutral model is formulated in terms of the source and destination communities, and in the present study, we apply Sloan model in pair-wise manner—each (of the 10 DT sites) acting as the source community paired with each (of the other 9 DT sites) as destination community (see Figure 1 for the illustration).

The stochastic neutrality analysis framework by Ning et al[20]

There was suggestion that the UNTB might over-estimate the true strength of neutral processes, we adopted Ning et al[20] normalized stochasticity ratio (NSR) as an alternative approach to gaging the "low bounds" of the stochasticity level. The principal foundation of Ning et al[20] mathematical framework is that deterministic processes should drive ecological communities more similar or dissimilar than null expectation. Ning et al[20] established a sophisticated procedure to implement a null model for quantifying stochasticity. A key metric in their framework was the utilization of Ružička[26] similarity metrics, which is a generalization of Jaccard binary similarity coefficient and is defined based on species abundance. Let C_{ij} represent the observed similarity between the i-th and j-th community,

$$C_{ij} = \frac{\sum_S \min(p_k^i, p_k^j)}{\sum_S \max(p_k^i, p_k^j)} \quad (1)$$

where S is the number of species, p_k^i and p_k^j are the relative abundance of k-th species in the i-th and j-th community.

Assume that there exist m local communities in a metacommunity, C_{ij} represents for the observed similarity between the i-th local community and the j-th local community in the metacommunity. E_{ij} represents for the null expected similarity between the i-th community and the j-th community in one simulated metacommunity. $\overline{E_{ij}}$ represents for the average of the null expected similarity between the i-th and the j-th communities in 1000 simulated metacommunities. Two possibilities exist in the assessment of the community stochasticity. One is that deterministic processes drive communities more similar, in which $C_{ij} > \overline{E_{ij}}$, and the stochasticity ratio (type A SR) is

$$SR_{ij}^A = \frac{\overline{E_{ij}}}{C_{ij}}. \quad (2)$$

Another possibility is that deterministic processes drive communities more dissimilar, in which $C_{ij} < \overline{E_{ij}}$, and the stochasticity ratio (type B SR) is

$$SR_{ij}^B = \frac{1-\overline{E_{ij}}}{1-C_{ij}}. \quad (3)$$

The stochasticity ratio in the whole metacommunity is

$$SR = \frac{\sum_{ij}^{n^A} SR_{ij}^A + \sum_{ij}^{n^B} SR_{ij}^B}{n^A + n^B}, \quad (4)$$

in which n^A represents for the number of the pair-wise similarities that are larger than null expectation, and n^B represents for the number of the pair-wise similarities that are less than null expectation. SR represents for the *strength of stochasticity* in the community assembly, and should range from 0% to 100%. If the community assembly is extremely deterministic without any stochasticity, then SR should be 0%; otherwise SR should be 100%. However, when expected stochasticity is very low, SR could overestimate stochasticity. To overcome this issue, SR should be normalized, and the normalized stochasticity ratio (NSR) exhibits higher precision than the SR and its exact definition and computational procedure are referred to Ning et al.[20] In the present article, we use Ning et al[20] normalized stochasticity ratio (NSR) to gain community-level insights on the niche-neutral continuum of human DT microbiomes.

Integrated with the OTU-level insights gained from Sloan near-neutral modeling, it is aimed to obtain educated guesses for the major parameters that can be used to quantitatively characterize the niche-neutral continuum of human DR microbiomes. Similar to that Sloan et al[18,19] near-neutral model that designates source and destination community, Ning et al[20] normalized stochasticity ratio (NSR) measures the stochasticity (stochastic drifts) between (pair-wise) 2 communities by measuring their similarity. In other words, the NSR is computed pair-wisely for the 10 DT microbiome sites, as illustrated in Figure 1.

While the previously introduced Sloan et al[18,19] model estimates the neutrality level (in percentage of neutral OTUs) or selection level (in percentages of positively or negatively selected OTUs) at the OTU-level, Ning et al[20] NSR assesses the stochasticity (≈neutrality) at the community (metacommunity) level. Putting together, both metrics, ranged between 0% and 100%, present a rational estimation of the niche-neutral continuum or spectrum from total deterministic selections (zero end) to completely stochastic drifts (one end) in the human DT microbiome.

Results and Discussion

OTU-level neutrality/selection analysis with Sloan et al[18,19] near-neutral model

We fitted Sloan et al[18,19] near-neutral model in pair-wise manner to the DT microbiomes, that is, building a Sloan model for each pair of the 10 DT microbiome sites, one site as source community and another site as destination (local) community (see Supplemental Table S1 and S2 for the detailed fitting results). Figure 2 illustrated the fitting of Sloan model to 2 pairs of DT microbiomes. Table 1 and Figure 3 further summarized the results on average basis for each DT site from Supplemental Table S1.

First, note that the Sloan model failed to fit the stool microbiome data, as judged by $R^2 < 0$, for this reason, we primarily rely on Ning et al[20] NSR metric to evaluate the "position" of stool on the niche-neutral continuum. However, the significantly lower NSR = 0.067 of stool (3-12 times lower than other 9 DT sites) seems to suggest that the actually parameters of Sloan model for stool are not without merits. Instead, the negative R^2 may have to do with some less satisfactory processing of the residual computation in the computational program used to fit Sloan model. In other words, we believe that the structure of Sloan model *per se* is solid. For this reason, we preserve the parameters of Sloan model fitted to Stool in Table 1 and Supplemental Tables S1 to S2. But we caution that for the stochasticity of stool samples (site), the inference should be primarily based on Ning et al[20] NSR. As a further caution, in all relevant tables and figures, we reported both versions of the aggregation metrics (mean, standard error), one with stool parameters included, and another excluded.

Overall but excluding the stool site, Table 1 shows that the stochastic neutral drifts, as measured by the proportion of neutral OTUs (46%), are slightly higher than that of the positive selection (38%), as represented by the proportion of positively selected OTUs. In addition, there are approximately 15% of OTUs that are negatively selected, and their abundances are smaller than the levels predicted by the neutral model. From

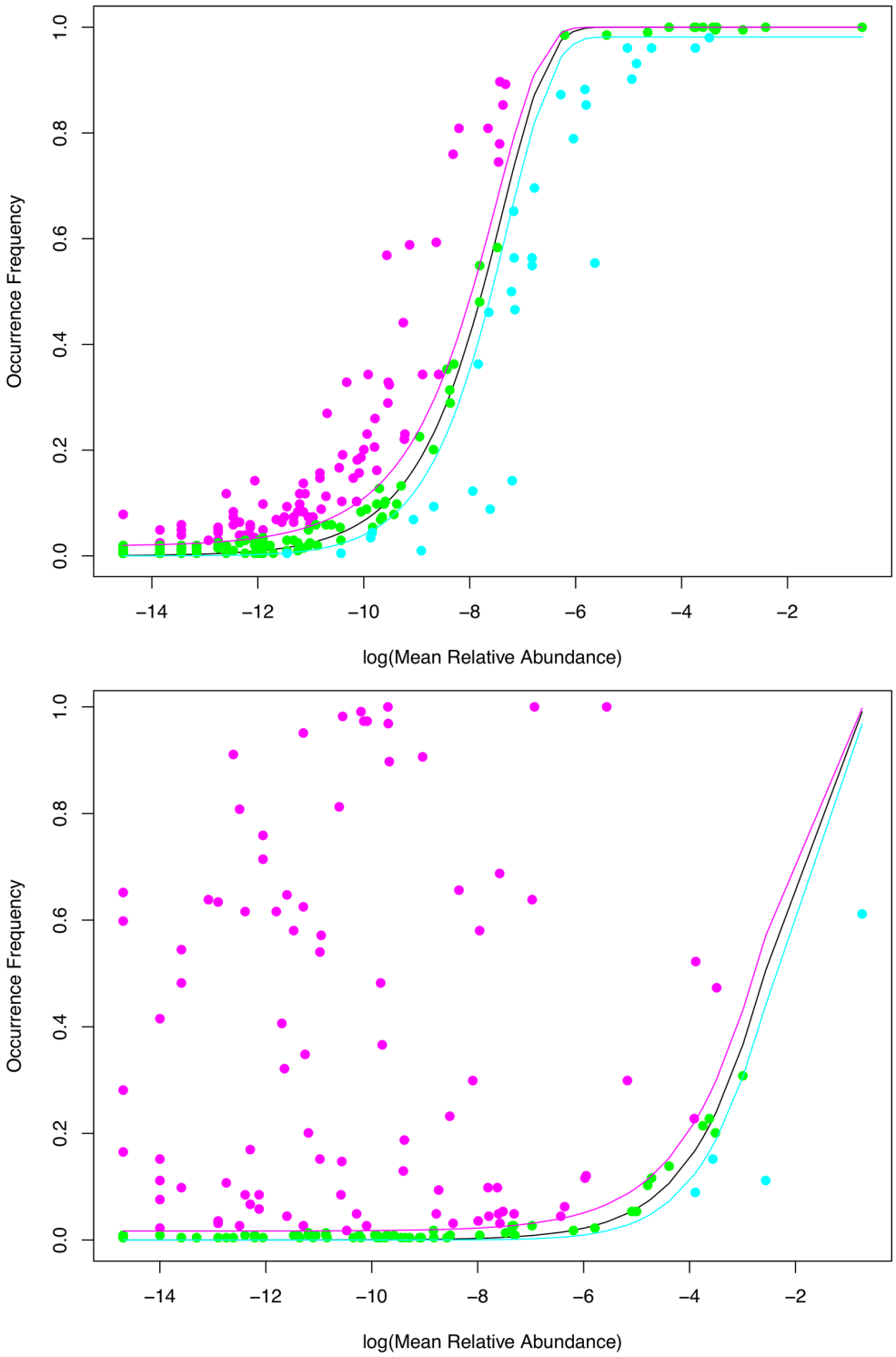

Figure 2. Fitting Sloan et al[18,19] near neutral model to the DT microbiome; the top graph for the BM-HP pair (Buccal mucosa vs Hard Palate), and bottom graph for BM-Stool (Buccal mucosa vs Stool) pair: (i) The 3 curves constitute the 95% confidence interval predicted with Sloan model; therefore, the pink circles representing for the species with their occurrence frequency > that of neutral species (ie, the positively selected species), the green circles for the neutral species, and the cyan for the species with their occurrence frequency < neutral species (the negatively selected species). (ii) The 2 graphs show contrastingly different distribution pattern of 3 categories of species; in the case of gut (stool) microbiome (the bottom), selection effects (such as diet effects from fully digested food) may be responsible for significantly large proportion (62%, pink circles) of positively selected species, while the proportion of neutral species (46%, the top graph, green circles) was much higher for the oral sites (Table 1). It should be cautioned that the fitting to Sloan model with stool microbiome data failed if judged by R^2, and the failure is also obvious in this figure, where positively selected OTUs scattered all over the places above the simulated neutral curve. We argue that even though the model fitting for stool microbiome failed, the proportions of the 3 categories of OTUs classified by Sloan model are still of important reference value, that is, nearly twice more positively selected OTUs, which is also consistent with the finding from Ning et al[20] NSR.

Table 1. Mean parameters of Sloan et al[18,19] near-neutral models fitted to the pair-wise DT microbiomes, summarized from Supplemental Table S1 (also refer to Supplemental Table S2 for additional information including AIC and BIC).

SOURCE COMMUNITY	IMMIGRATION PROBABILITY (*m*)	R^2	TOTAL NUMBER OF SPECIES	PERCENTAGE OF SPECIES BELOW NEUTRAL (%)	PERCENTAGE OF NEUTRAL SPECIES (%)	PERCENTAGE OF SPECIES ABOVE NEUTRAL (%)
BM	0.053	0.691	212	15.8	51.5	32.6
HP	0.025	0.638	224.2	16.4	50.2	33.3
KG	0.255	0.494	183.1	11.8	43.8	44.4
PT	0.015	0.688	210.1	12.5	47.0	40.5
Sal	0.018	0.712	216.3	14.3	49.7	36.0
SubP	0.011	0.604	202.1	15.1	41.6	43.3
SupP	0.027	0.544	185	15.2	35.7	49.1
TD	0.064	0.650	168.1	16.1	36.9	47.0
Th	0.012	0.650	229.9	14.9	51.9	33.2
Stool	0.002	−0.440	162.5	11.0	27.2	61.8
Summary statistics of the Sloan model parameters computed from all pairs of DT sites						
Mean	0.048	0.523	199.33	14.31	43.55	42.12
Std. Err.	0.024	0.109	7.398	0.597	2.590	2.895
Summary statistics of the Sloan model parameters computed from all pairs except for the Gut (Stool) site						
Mean	0.059	0.751	208.580	15.722	46.000	38.256
Std. err.	0.029	0.026	7.206	0.578	2.182	2.269

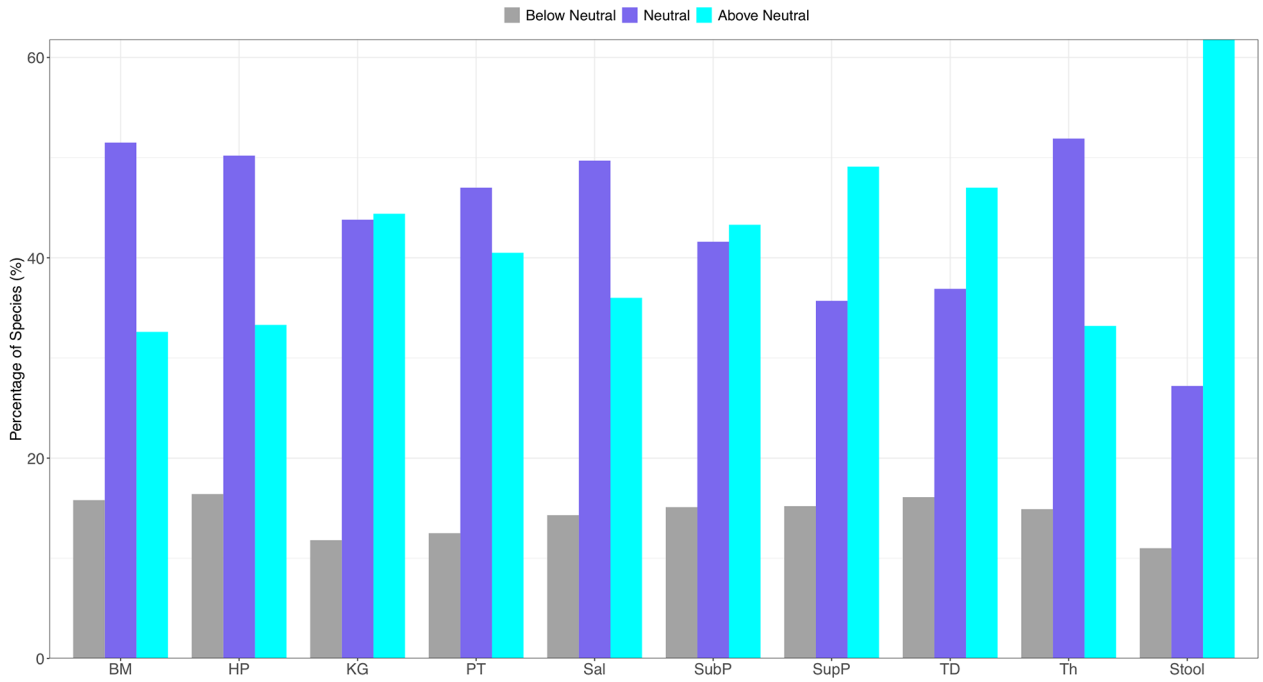

Figure 3. The percentages of neutral, above neutral (positively selected), and below neutral (negatively selected) species at each of the 10 DT sites: the gut (stool) showing significantly higher number of positively selected (above neutral) species, while other sites (most are oral and throat showing more neutral species (approximately 46 on average) (see Table 1 for the details).

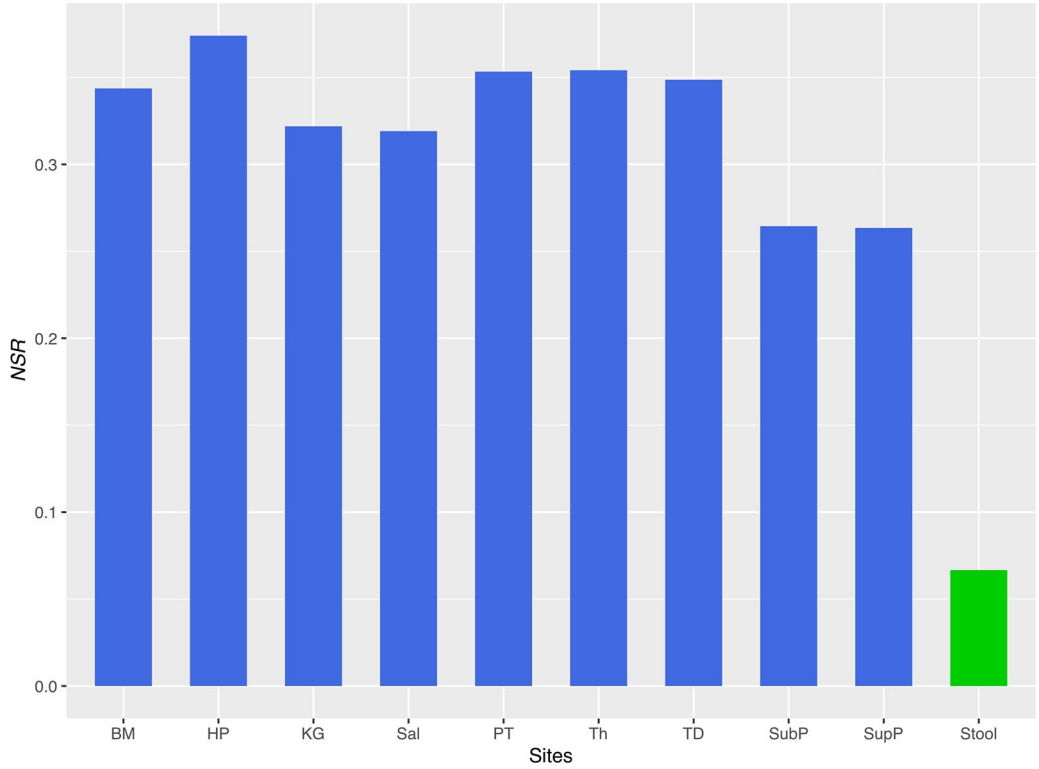

Figure 4. The average NSR (normalized stochasticity ratio) for each of the 10 DT sites, averaged from pairing with other 9 DT sites.

the perspective of niche-neutral continuum, one can say that both stochastic neutral drifts (such as stochasticity in birth, death and immigration) and deterministic selection forces (such as site-specific habitats) appear equally important in shaping the structure and dynamics of DT microbiomes. It should be reiterated that gut (stool) seems to have significantly lower neutrality and higher positive selection than other DT sites, as further evidenced by the 3 to 12 times lower of NSR below (Table 2). As rightly pointed out by an anonymous reviewer, the stool microbiota mostly reflects that of the colonic area, with an extreme high cell density and low relative (per cell) concentration of nutritional resources, which may lead to strong interbacterial competition in the form of positive selection. While, these insights are obviously from the OTU-level analysis, in the next section, we expose the results from community-level stochasticity analysis.

Community-level stochasticity analysis of the niche-neutral continuum with the NSR

Table 2 listed the results from using Ning et al[20] normalized stochasticity ratio (NSR). The NSR of DT microbiomes ranged between 0.018 and 0.575, with gut (stool) microbiome exhibiting a significantly lower NSR value of 0.067, approximately one-fifth of the NSR of other sites. This may be explained by the fact that food is mostly digested in the gut (stool) habitat and is likely to exert more specialized (selection) influences on microbes, while food in other DT sites (most are oral and throat sites) we analyzed have less specialized influences on microbes. Consequently, in the gut habitat, selection is stronger and neutral drift is weaker than in the other sites.

Figure 4 displayed the average NSR for each of the 10 DT sites, averaged from pairing with 9 other DT microbiome sites. The NSR of DT microbiomes suggest that the upper section of DT exhibits significant portion of stochasticity or neutrality (approximately 0.36), while the stool exhibits nearly negligible stochasticity (0.067). These findings are consistent with those from previous OTU-level analyses. Therefore, the community-level stochasticity framework analysis seems to depict a picture of niche-neutral continuum along the DT: the upper DT section appears to fall on the neutral side of the continuum, while the lower terminal side of DT (stool) appears to fall on the niche side of the continuum. We postulate that this pattern (or spectrum) of DT niche-neutral continuum has to do with the nature of food in DT. In the upper section of the DT, the food is largely raw and exerts relatively weak selection forces to the microbes, and the food is more thoroughly digested and exerts strong and specialized (niche) selection forces to the microbes in the lower section. Particularly in the gut, the previously mentioned eco-active intestinal chemosphere[24] may lead to strong selection effects. Note that the term weak (or strong) selection we use in this article means that neutral drifts are relatively strong (or weak), possibly due to relatively stronger selection force in the lower DT section, especially from

Table 2. The average similarity (S) and normalized stochasticity ratio (NSR) for each pair among the 10 DT sites, summarized from Supplemental Table S3.

SITE ID	SIMILARITY (*S*)	NORMALIZED STOCHASTICITY RATIO (NSR)
BM	0.717	0.344
HP	0.730	0.374
KG	0.727	0.322
Sal	0.713	0.319
PT	0.722	0.353
Th	0.725	0.354
TD	0.724	0.349
SubP	0.693	0.264
SupP	0.694	0.263
Stool	0.819	0.067
Summary Statistics of all pairs of DT sites		
Mean	0.726	0.301
Std. Err.	0.008	0.021
Range	0.65-0.829	0.018-0.575
Summary Statistics of all pairs of DT sites except for the Stool site		
Mean	0.703	0.360
Std. Err.	0.005	0.015
Range	0.65-0.791	0.208-0.575

intestinal chemosphere. Our usage of weak selection is related but different from Wu et al[27] usage in the context of evolutionary game theory, in which weak selection means that the game has only a small influence on evolutionary dynamics.

Quantification of niche differentiations

In previous subsections, we envisioned and interpreted the niche-neutral continuum of the human DT microbiomes from species and community level analyses. Here we further analyze the niche differentiations by drawing heatmaps with the immigration probability (m) from Sloan et al[18,19] near-neutral model and NSR from Ning et al[20] stochasticity framework, respectively. Figure 5 illustrated the heatmap based on Sloan et al[18,19] near-neutral model, and Figure 6 illustrated the heatmap based on Ning et al[20] NSR framework. Both approaches cluster the 10 DT microbiome sites as exactly same 5 groups, as illustrated in Figures 5 and 6. Four of the 5 site types (niches) actually correspond to the 4 niches previously discovered by Segata et al,[2] including: (i) Stool; (ii) sub-gingival plaques (SubP) and supra-gingival plaques (SupP); (iii) tongue dorsum (TD), throat (TH), palatine tonsils (PT), and saliva (Sal); and (iv) hard palate (HP) and buccal mucosa (BM), and *keratinized gingiva* (KG). However, the site of *keratinized gingiva* (KG), which was originally classified into the same niche with HP and BM, may represent a potentially distinct niche *candidate*. In fact, the site KG is dissimilar with any of the other 9 sites, as illustrated in Figures 5 and 6. The stand-alone nature of KG is even more obvious in Figure 7, which compared the migration probability (m) of 10 DT microbiome sites. The migration probability ($m = 0.255$) of KG site is 4 to 128 times that of the other 9 sites (0.002-0.064). This suggests that KG might be a transitional site for microbial migrations, a possibly fifth niche previously ignored, while Segata et al[2] original study only revealed 4 distinctive site types (niches).

Existing studies (eg,[28,29]) may offer some clues for explaining the distinctiveness of the KG site. There have been many studies on the role of KG in marginal periodontal behavior. Some studies suggested that a minimum of 2.0 mm of KG width is needed to maintain marginal periodontal health, and others suggested that KG width is negligible if excellent oral hygiene is maintained.[28] KG is found to play an important role in marginal periodontal homeostasis, and specialist microbes of KG site should have far reaching implications to dental health.[28,29] Most recently, Chen and Ma[21] also classified, through multi-site neutral and niche-neutral hybrid models, the KG as a separate niche, distinctively different from hard palate (HP) and buccal mucosa (BM).

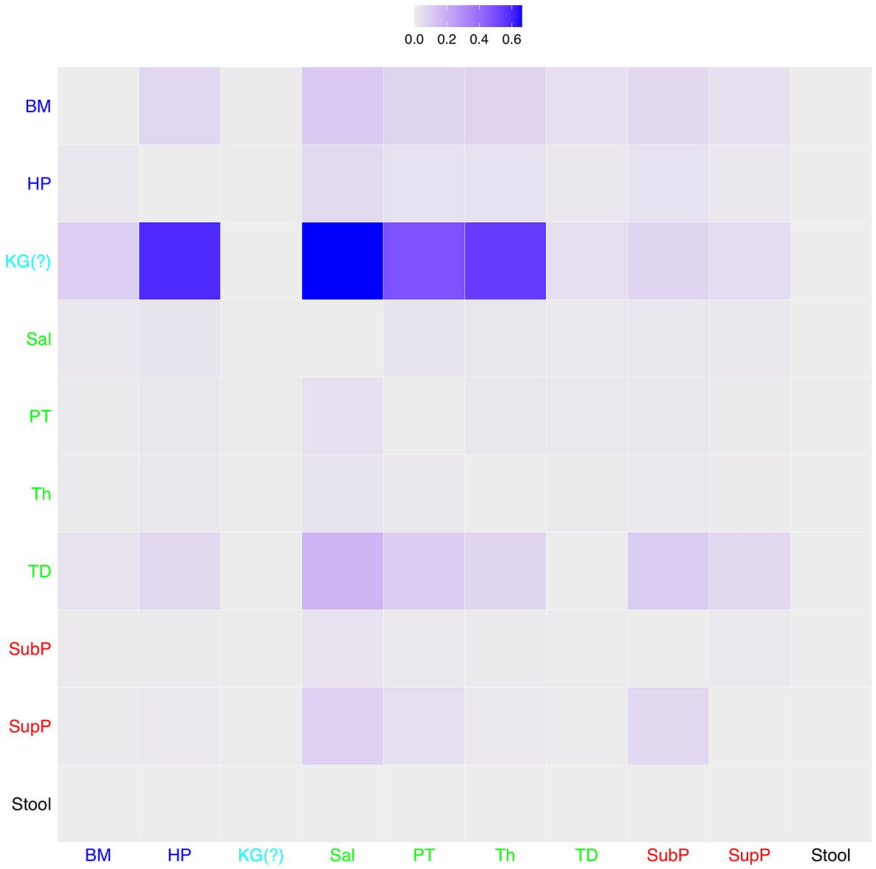

Figure 5. A heatmap showing the *m* (mean migration probability) of the 10 sites: the larger the *m*-value is (the deeper the color in the heatmap is), the higher the similarity between the DT sites (see Supplemental Table S1 for detailed results of other Sloan model parameter). The site KG (keratinized gingival), together with the 4 niches (the 4 types of DT microbiota, ie, "BM-HP-KG," "Sal-PT-Th-TD," "SubP-SupP," and "Stool," colored differently, and originally discovered by Segata et al[2]) emerged as 5 clusters here. The KG site was originally classified into the "BM-HP-KG" type by Segata et al,[2] but it is obviously different from the other 4 clusters (also see Figure 7), likely more deeply shaped by deterministic selection force (also *refer* to Figure 7 for its exceptionally high immigration probability).

Discussion

In summary, whether it is the niche-neutral continuum hypothesis or Vellend[30] and Hanson et al[31] 4-processes (drift, selection, dispersal, and speciation) synthesis, a key to deciphering their underlying mechanisms is to estimate the stochastic neutral drifts, which can be performed with Sloan et al[18,19] near-neutral modeling at the species or OTU level, or Ning et al[20] normalized stochasticity ratio (NSR) at the community level. Sloan model offers OTU-level insights that both stochastic neutral drifts and deterministic positive selection play nearly equal roles, except for the stool microbiome site, in shaping the assembly of DT microbiomes. The community-level stochasticity analysis with Ning et al[20] NSR analysis generated the similar findings as the OTU-level analysis did. Although Sloan model failed to fit the stool microbiome data, Ning et al[20] NSR revealed that the stool microbiome seem to have experienced much stronger deterministic selections, likely 3 to 12 times stronger than the other sites do. We postulate that the strong positive selection occurring in stool microbiome should be attributed to the mostly digested food that exerts more specialized (niche differentiation effects) selections to stool microbes. One additional finding of this study is a possibly fifth niche (KG), as demonstrated by its exceptionally high migration probability, which suggests that KG might be a particularly active transitional site for microbial dispersal along the DT, a possibly fifth niche previously ignored. This new proposal (hypothesis for a fifth niche) seems to support the distinctive characteristics of KG previously discussed by Lagos et al[28] and Mark Welch et al.[29]

Compared with the microbiomes of upper DT sites (mouth and throat), the microbiomes of lower DT sites (intestine and stomach) are likely shaped by much stronger selection from the intestinal (eco-active) chemosphere, as briefly discussed previously. The concept of intestinal (eco-active) chemosphere[24] refers to the ensemble of chemical molecules in the lumen and on the surface of gut, namely, the chemical substances that promote or inhibit bacterial growth. Those chemical molecules of intestinal chemosphere can originate from nutrient uptake, or be produced by the host and their intestinal microbiome.

We postulate that the assembly and diversity maintenance of DT microbiomes could be deeply influenced by the intestinal chemosphere. The intestinal chemosphere is part of the "invironment" of the DT microbiomes—"a shared space where the

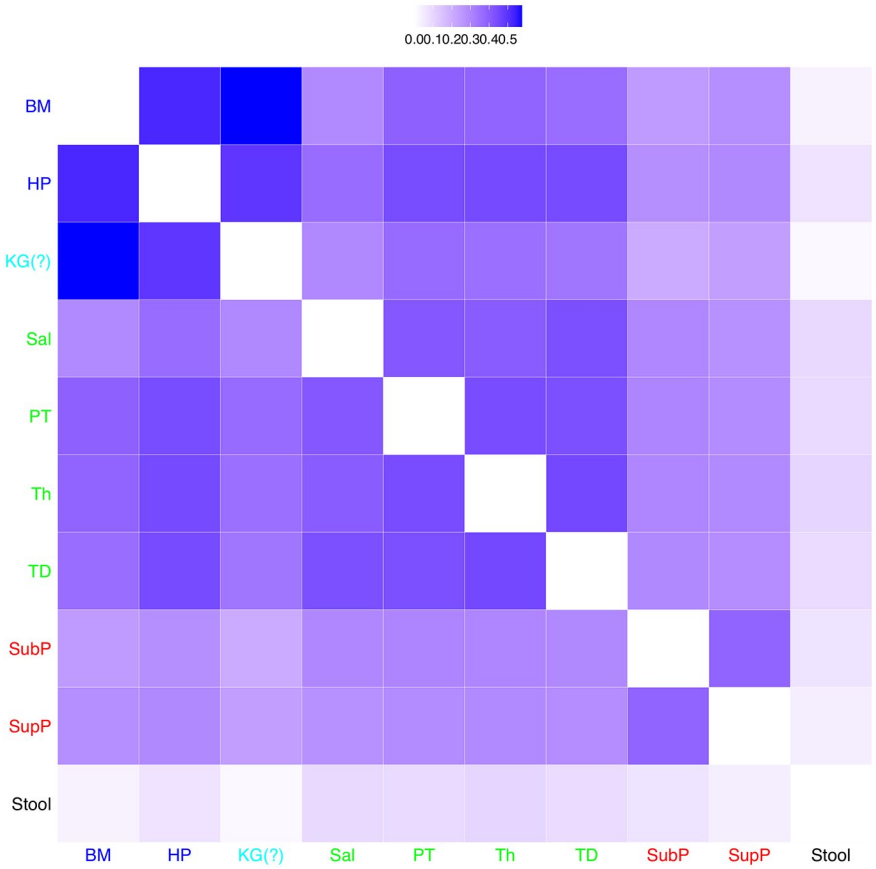

Figure 6. A heatmap showing the pair-wise NSR value between the 10 DT sites: the larger the NSR is (the deeper the color in the heatmap is), the higher the similarity between the DT sites (see Table 2 and Supplemental Table S3 for detailed information).

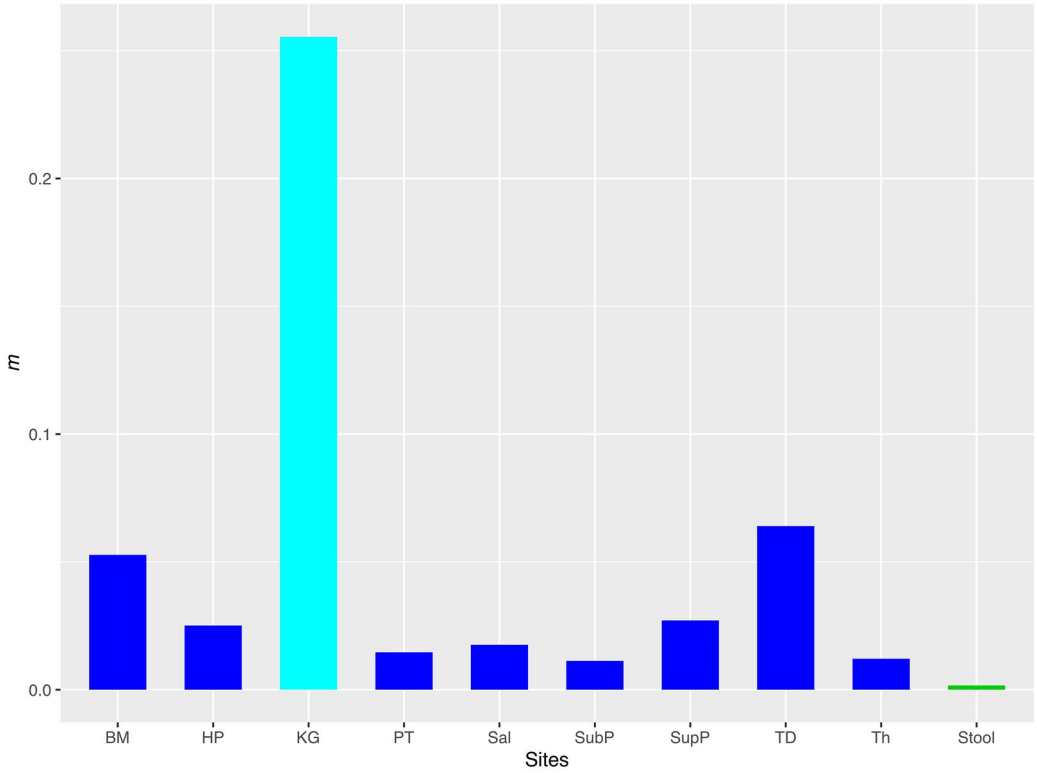

Figure 7. The mean migration probability (*m*) of each site computed from Supplemental Table S1. The KG site was originally classified into the "BM-HP-KG" type by Segata et al,[2] but it is obviously rather unique, likely more deeply shaped by deterministic selection force than the neighboring sites. The migration probability (*m*) is 4 to 128 times higher than the other 9 sites.

interior and exterior of the organism merge."[24,32] The term "invironment" was coined by Baquero[32] to reflect somewhat uniqueness of DT microbiome environment, which is both "open" and "closed." It is open, because the DT is partially influenced by factors external to the host and microbiomes such as food, swallowed environmental microorganisms, and abiotic environmental factors. It can be considered as closed given that much of the intestinal mucosal cells and lumen fluids are not exposed to the outside of the host. Therefore, the future studies on the niche-neutral continuum of DT microbiomes should certainly take the intestinal chemosphere into consideration.

There are limitations with our study. The first is that the middle section (stomach and small intestines) of the DT was not sampled. In addition, some of the sites may also be considered as habitats of the respiratory microbiomes, and in most existing studies, the intestinal microbiome starts in the esophagus, not in the mouth. Strictly speaking, DT should be considered as the tract where solid and liquid food are in touch with human mucosa to be processed as nutrients, and overlapping respiratory tract is the tract where inhaled air is in touch with human mucosa.[33] Furthermore, the delineation of 4[2] or proposed 5 niches may simply be the "checkpoints" on the niche-neutral continuum, likely influenced by the sampling scheme adopted by Segata et al.[2] A second limitation of this study is to do with the computational nature of this study, which depends on statistical inferences and simulations to infer ecological insights from species abundance distributions (SAD) in the human DT microbiome. The SAD data were from sequencing microbiome samples, and no manipulative experiments (which are generally infeasible for healthy humans for ethic reasons) data are available. Therefore, any of the mechanistic claims (parameters) in this study should be treated as educated guesses, rather than precise numbers or mechanisms. A third limitation is that our study, like most existing studies, failed to consider the compartmentalization of DT, which can be considered as one of the most compartmentalized organs in our body, but our treatment of the DT sites was essentially in "linear" manner. Baquero and Negri[34] discusses the influences of compartmentalization of our body parts on the development of bacterial resistance to antibiotics, which is of obvious significance for clinical medicine. Similarly, it is likely that different compartments (eg, stomach) may have differentiated into different niches. This limitation also implicates our usage of the niche-neutral continuum along the DT, which implies a "linear" spectrum in the level of neutrality. In this sense, our usage of the term "continuum" should be considered as an analogy. Of course, the usage of continuum is not related to its usage in real analysis (mathematics), and is pure ecological sense.

Acknowledgements

We appreciate the computational support from Dr. LW Li and Dr. Wendy Li for illustration help, both from the Computational Biology and Medical Ecology lab of the Chinese Academy of Sciences. We are particularly indebted to two anonymous expert reviewers for their insightful comments, suggestions, and advice.

Authors' Contributions

ZSM designed, performed the analysis and wrote the paper. HJC performed data analysis. All authors approved the submission.

Additional Statements

The study does not involve any physical or biological experiments, other than computational analyses. No experimental protocols were used. No informed consent is applicable since no human or animal subjects were used for any experiment.

Data Availability Statement

The DT dataset is available in Segata et al[2] supplementary materials. The computational code for Sloan near-neutral model was published in Sloan et al.[18,19] The computational code for Ning stochasticity framework was published in Ning et al.[20]

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. HMP Consortium. A framework for human microbiome research. *Nature*. 2012; 486(7402):215.
2. Segata N, Haake SK, Mannon P, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol*. 2012;13:R42.
3. Ma ZS. Niche-neutral theoretic approach to mechanisms underlying the biodiversity and biogeography of human microbiomes. *Evol Appl*. 2021;14:322-334.
4. Deines P, Hammerschmidt K, Bosch TC. Exploring the niche concept in a simple metaorganism. *Front Microbiol*. 2020;11:633429.
5. Hubbell SP. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press; 2001.
6. McGill BJ. A test of the unified neutral theory of biodiversity. *Nature*. 2003; 422:881-885.
7. McGill BJ, Maurer BA, Weiser MD. Empirical evaluation of neutral theory. *Ecology*. 2006; 87:1411–1423.
8. Purves W, Drew W, Pacala, Stephen. Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. In: Burslem D, Pinard M, Hartley S, eds. editors *Biotic Interactions in the Tropics: Their Role in the Maintenance of Species Diversity*. Cambridge University Press; 2005:108-138.
9. Alonso D, Etienne R, Mckane A. The merits of neutral theory. *Trends Ecol Evol*. 2006;21:451-457.
10. McGill BJ, Maurer BA, Weiser MD. Empirical evaluation of neutral theory. *Ecology*. 2006;87:1411-1423.
11. Kadmon R, Allouche O. Integrating the effects of area, isolation, and habitat heterogeneity on species diversity: A unification of island biogeography and niche theory. *Am Nat*. 2007;170:443-454.
12. Chisholm RA, Pacala SW. Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *PNAS*. 2010;107:15821-15825.
13. Rosindell J, Hubbell SP, Etienne RS. The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol Evol*. 2011;26:340-348.
14. Tang J, Zhou S. Hybrid niche-neutral models outperform an otherwise equivalent neutral model for fitting coral reef data. *J Theor Biol*. 2013;317:212-218.
15. Matthews TJ, Whittaker RJ. Neutral theory and the species abundance distribution: recent developments and prospects for unifying niche and neutral perspectives. *Evol Ecol*. 2014;4:2263-2277.
16. Al Hammal O, Alonso D, Etienne RS, Cornell SJ. When can species abundance data reveal non-neutrality? *PLoS Comput Biol*. 2015;11:e1004134.
17. Gilbert B, Levine JM. Ecological drift and the distribution of species diversity. *Proc Biol Sci*. 2017;284:20170507.

18. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol*. 2006;8:732-740.
19. Sloan WT, Woodcock S, Lunn M, Head IM, Curtis TP. Modeling taxa-abundance distributions in microbial communities using environmental sequence data. *Microb Ecol*. 2007;53:443-455.
20. Ning D, Deng Y, Tiedje JM, Zhou J. A general framework for quantitatively assessing ecological stochasticity. *PNAS*. 2019;116:16892-16898.
21. Chen HD, Ma ZS. Niche-neutral continuum seems to explain the global niche differentiation and local drift of the human digestive tract microbiome. *Front Microbiol*. 2022;13:912240.
22. Harris K, Parsons TL, Ijaz UZ, Lahti L, Holmes I, Quince C. Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical Dirichlet process. *Proc IEEE*. 2017;105:516-529.
23. Baquero F, Coque TM, Galán JC, Martinez JL. The origin of niches and species in the bacterial world. *Front Microbiol*. 2021;12:657986.
24. Baquero F, Lanza VF, Baquero MR, Del Campo R, Bravo-Vázquez DA. Microcins in Enterobacteriaceae: peptide antimicrobials in the eco-active intestinal chemosphere. *Front Microbiol*. 2019;10:2261.
25. Burns AR, Stephens WZ, Stagaman K, et al. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J*. 2016;10:655-664.
26. Ružička M. *Anwendung Mathematisch-statisticher Methoden in der Geobotanik (Synthetische Bearbeitung von Aufnahmen)*. Biologia; 1958.
27. Wu B, Altrock PM, Wang L, Traulsen A. Universality of weak selection. *Phys Rev E*. 2010;82:046106.
28. Lagos ML, Sant'ana AC, Greghi SL, Passanezi E. Keratinized gingiva determines a homeostatic behavior of gingival sulcus through transudation of gingival crevice fluid. *Indent*. 2011;2011:953135.
29. Mark Welch JL, Dewhirst FE, Borisy GG. Biogeography of the oral microbiome: the site-Specialist Hypothesis. *Annu Rev Microbiol*. 2019;73:335-358.
30. Vellend M. Conceptual synthesis in community ecology. *Q Rev Biol*. 2010;85:183-206.
31. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JB. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol*. 2012;10:497-506.
32. Baquero F. Metagenomic epidemiology: a public health need for the control of antimicrobial resistance. *Clin Microbiol Infect*. 2012;18 Suppl 4:67-73.
33. Pérez-Cobas AE, Baquero F, de Pablo R, Soriano MC, Coque TM. Altered ecology of the respiratory tract microbiome and nosocomial pneumonia. *Front Microbiol*. 2021;12:709421.
34. Baquero F, Negri MC. Challenges: selective compartments for resistant microorganisms in antibiotic gradients. *Bioessays*. 1997;19:731-736.