



OPEN ACCESS

EDITED BY

Francesco Dal Grande,
University of Padua,
Italy

REVIEWED BY

Shan Sun,
University of North Carolina at Charlotte,
United States
Shi Huang,
The University of Hong Kong,
Hong Kong SAR, China

*CORRESPONDENCE

Qichao Tu
tuqichao@sdu.edu.cn

SPECIALTY SECTION

This article was submitted to
Evolutionary and Genomic Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 04 August 2022

ACCEPTED 15 September 2022

PUBLISHED 06 October 2022

CITATION

Ma K and Tu Q (2022) Random sampling
associated with microbial profiling leads to
overestimated stochasticity inference in
community assembly.
Front. Microbiol. 13:1011269.
doi: 10.3389/fmicb.2022.1011269

COPYRIGHT

© 2022 Ma and Tu. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC
BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Random sampling associated with microbial profiling leads to overestimated stochasticity inference in community assembly

Kai Ma¹ and Qichao Tu^{1,2*}

¹Institute of Marine Science and Technology, Shandong University, Qingdao, China, ²Joint Lab for Ocean Research and Education at Dalhousie University, Shandong University and Xiamen University, Qingdao, China

Revealing the mechanisms governing the complex community assembly over space and time is a central issue in ecology. Null models have been developed to quantitatively disentangle the relative importance of deterministic vs. stochastic processes in structuring the compositional variations of biological communities. Similar approaches have been recently extended to the field of microbial ecology. However, the profiling of highly diverse biological communities (e.g., microbial communities) is severely influenced by random sampling issues, leading to undersampled community profiles and overestimated β -diversity, which may further affect stochasticity inference in community assembly. By implementing simulated datasets, this study demonstrate that microbial stochasticity inference is also affected due to random sampling issues associated with microbial profiling. The effects on microbial stochasticity inference for the whole community and the abundant subcommunities were different using different randomization methods in generating null communities. The stochasticity of rare subcommunities, however, was persistently overestimated irrespective of which randomization method was used. Comparatively, the stochastic ratio approach was more sensitive to random sampling issues, whereas the Raup–Crick metric was more affected by randomization methods. As more studies begin to focus on the mechanisms governing abundant and rare subcommunities, we urge cautions be taken for microbial stochasticity inference based on β -diversity, especially for rare subcommunities. Randomization methods to generate null communities shall also be carefully selected. When necessary, the cutoff used for judging the relative importance of deterministic vs. stochastic processes shall be redefined.

KEYWORDS

random sampling, β -diversity, microbial community, stochasticity, null models, Raup–Crick metric, sequencing depth

Introduction

Revealing the mechanisms governing the complex community assembly over space and time is a central issue in ecology. Two distinct types of theories, including the niche theory and the neutral theory (Vandermeer, 1972; Hubbell, 2001), have been developed to explain the compositional variations of biological communities. Historically, the niche theory has gained great success in explaining the dynamic changes in community composition in various ecosystems (Harpole and Tilman, 2007; O'Malley, 2007; Holt, 2009; Kylafis and Loreau, 2011). However, the existence of highly diverse ecosystems such as rainforest, in which many organisms coexist in a same ecological niche (Hubbell, 1979; Scheffer and van Nes, 2006), challenges the throne of niche theory in community ecology. To solve such issues, Hubbell (2001, 2011) proposed the neutral theory, by which many challenges in community ecology can be well resolved. Until now, a general consensus has been reached by ecologists that both deterministic (niche theory) and stochastic (neutral theory) processes shape the assembly of biological communities, but their relative importance may differ in different ecosystems (Dumbrell et al., 2010; Ofițeru et al., 2010; Chase and Myers, 2011; Stegen et al., 2012; Fisher and Mehta, 2014). Interestingly, recent studies show that sampling scale could be an important factor affecting the relative importance inference of determinism vs. stochasticism in shaping community assembly (Chase, 2014; Correa-Metrio et al., 2014).

Similar issues have been recently recurred in microbial community ecology. Over the last decade, our understanding regarding the complex microbial community assembly has been revisited. For many years, the niche theory has dominated the field with studies mainly focusing on environmental factors that structure the diversity and composition of microbial communities (Fierer and Jackson, 2006; Martiny et al., 2006; Lozupone and Knight, 2007; Fierer et al., 2009; Fuhrman, 2009; Caporaso et al., 2011; Freitas et al., 2012; Thompson et al., 2017; Oliverio et al., 2020). Such efforts can date back as early as to 1930s when Baas Becking proposed the famous hypothesis “Everything is everywhere, but, the environment selects” (Baas-Becking, 1934; Wit and Bouvier, 2006). Important progresses have been made toward our understanding of the relationship between environmental factors and microbial communities. For instance, pH and temperature are found as important factors shaping the diversity and composition of soil microbial communities at large scales (Griffiths et al., 2011; Shen et al., 2013; Sunagawa et al., 2015; Tu et al., 2016; Zhou et al., 2016; Jiao and Lu, 2020; Liang et al., 2020; Xu et al., 2020). Recent studies also demonstrate that both deterministic and stochastic processes play critical roles in structuring the immense microbe world (Zhou et al., 2014; Dini-Andreote et al., 2015; Xue et al., 2018; Nyirabuhoro et al., 2020), and the question to resolve is which process is relatively more important (Antwis et al., 2017; Zhou and Ning, 2017). More recently, studies show that organism size (Farjalla et al., 2012; Wu et al., 2018; Luan et al., 2020) and spatial scale (Shi et al., 2018;

Zhang et al., 2020b; Song et al., 2022) are also critical factors influencing the relative importance of deterministic and stochastic processes in structuring microbial communities.

Microbial communities are substantially different from microbial communities regarding the diversity and the role of rare taxa. Typical microbial communities are composed by a small set of abundant taxa and an extremely long tail of rare taxa (Lynch and Neufeld, 2015). The abundant subcommunity usually occupies <20% of the total richness, but comprises >80% in relative abundance (Sogin et al., 2006; Lynch and Neufeld, 2015; Zhang et al., 2020a). Notably, studies suggest that the abundant taxa are usually abundant, whereas the rare taxa are persistently rare (Galand et al., 2009). Such scenario also holds true when looking at more systematic microbial community data generated by the Earth Microbiome Project (EMP; Gilbert et al., 2014), the Human Microbiome Project (HMP; Turnbaugh et al., 2007), and the TARA Oceans Expedition (Pesant et al., 2015). Although low in relative abundance, recent studies suggest that the rare subcommunities execute nonnegligible ecosystem functions in the environment (Lyons and Schwartz, 2001; Lyons et al., 2005; Mouillot et al., 2013). For such reasons, efforts have been made to disentangle the underlying ecological mechanisms structuring rare subcommunities (Jia et al., 2018; Mo et al., 2018; Zhang et al., 2018; Jiao and Lu, 2020). Although carried out in different ecosystems, these studies suggest that the abundant and rare subcommunities are structured by different mechanisms (Jia et al., 2018). For instance, the rare subcommunities in subtropical ecosystems are more structured by stochastic processes than abundant subcommunities (Mo et al., 2018; Xue et al., 2018). Similar patterns are also observed for microbial communities in oil-contaminated soils (Jiao et al., 2017). While in the Qinghai-Tibet Plateau wetland ecosystem, it is found that variable selection (deterministic process) governs the community assembly of rare bacteria, whereas dispersal limitation (stochastic process) dominates community assembly of abundant bacteria (Wan et al., 2021).

Notably, the profiling of microbial communities is severely affected by random sampling issues, even using high-throughput sequencing approaches (Zhou et al., 2011, 2013; Zhan et al., 2014b; Tu, 2020). Random sampling issues are associated with each step the microbial communities are profiled, including sample collection, DNA extraction, library construction, amplification, sequencing, and subsequent rarefaction to a same sequencing depth (SeqD). This is mainly caused by the tiny size and high diversity of microbial communities in nature, as well as the limitations of current technologies that complete capturing every single microorganism is not feasible. As a result, only a small portion of the microorganisms in the collected samples are analyzed, leading to undersampled microbial profiles. Specifically, each gram of soil contains as high as 10^4 prokaryotic species and 10^8 organisms (Whitman et al., 1998; Torsvik and Øvreås, 2002; Daniel, 2005), while <100,000 sequences are usually captured for each sample. This number goes much lower after data processing

such as quality control and random subsampling/rarefaction to a same SeqD.

In this study, we investigated how microbial stochasticity inference was affected by such undersampled microbial profiles using simulated datasets. Previous studies suggest that random sampling issues associated with microbial profiling lead to overestimated β -diversity (Zhou et al., 2011, 2013; Zhan et al., 2014b; Tu, 2020). And, the effects of random sampling on abundant and rare subcommunities were dramatically different (Zhan et al., 2014a). Because microbial community stochasticity is usually inferred by comparing the observed β -diversity with null expectations, the overestimated β -diversity may lead to more similar/dissimilar patterns with null expectations. Therefore, we expected that microbial stochasticity may also be strongly affected, especially for rare subcommunities. Such effects may differ by different randomization methods generating null communities. By implementing well-controlled simulated datasets, the current study demonstrated solid evidence showing overestimated microbial stochasticity due to random sampling issues associated with microbial profiling, especially for rare subcommunities. Such overestimation eased with increasing SeqD, but could not be eliminated with current sequencing efforts. We therefore urge cautions be made for microbial stochasticity inference using null models.

Materials and methods

Methodological framework

A framework was developed to investigate the effects of random sampling issues associated with microbial community profiling on community stochasticity inference (Figure 1). In order to precisely quantify how microbial stochasticity was affected, simulated datasets were constructed and used in this study. First, pseudo seed communities containing 10^4 microbial taxa and 10^8 organisms were created. Based on the pseudo seed communities, seed communities with different levels of β -diversity were generated. Then, a seed metacommunity was formed by randomly selecting one of the each seed communities with different diversities and merging them. Second, mock (meta)communities were generated by random subsampling select numbers of organisms from the seed (meta) communities, representing the microbial communities obtained in typical microbial ecological studies. Multiple sets of mock (meta) communities with different organism numbers were generated in order to investigate whether increasing SeqD would eliminate the effects of random sampling issues. Third, microbial profiles were generated for both seed and mock metacommunities, based on which microbial community stochasticity was calculated. The community stochasticity for the seed and the mock metacommunities was comparatively analyzed, with the differences representing the effects of random sampling issues on microbial stochasticity inference.

Two different types of stochasticity analyses methods, including the stochastic ratio (Zhou et al., 2014; Zhou and Ning, 2017; Guo et al., 2018) and the Raup–Crick (RC) metric (Raup and Crick, 1979; Chase et al., 2011; Stegen et al., 2013, 2015), were employed here to evaluate how random sampling affected stochasticity inference.

Seed and mock community construction

A total of 15 pseudo seed communities were constructed following lognormal distributions, which is the species abundance distribution (SAD) model followed by most microbial communities in both natural and artificial ecosystems (Shoemaker et al., 2017). Each seed community was composed by 10^4 taxa and 10^8 organisms, representing the basic microbial diversity in per unit environmental samples (e.g., soil; Whitman et al., 1998). A select number (0 to 100%) of taxa in the seed communities were renamed as new taxa and/or randomly shuffled, mimicking community assembly processes such as drift and dispersal limitation. As a result, seed communities with different β -diversity (Bray–Curtis dissimilarity) were generated, and further seed metacommunities were obtained (Supplementary Table 1). Mock (meta)communities were then generated by random subsampling a select number (5×10^3 – 2×10^5) of organisms from the seed (meta) communities, representing microbial communities under different SeqD. Two major parameters associated with lognormal distribution, including “meanlog” and “sdlog,” were assessed here. The seed metacommunities were found with “meanlog” of 6.80 ± 0.03 and “sdlog” of 2.20 ± 0.00 , whereas the values for mock metacommunities (e.g., SeqD = 30,000) were, respectively, 1.00 ± 0.00 and 1.14 ± 0.00 (Supplementary Table 2). These values were comparable to what have been observed for microbial communities in different studies (Supplementary Table 2), such as the Earth Microbiome Project (EMP; Gilbert et al., 2014), the TARA Oceans expedition (Pesant et al., 2015) and the Human Microbiome Project (HMP; Turnbaugh et al., 2007). R packages “mobsim” (May et al., 2018) and a developed R script rarefy_vt.R were, respectively, used for seed community and mock community constructions.

Defining abundant and rare taxa

No standard is currently available for the definition of abundant and rare microbial taxa in complex communities. Different criteria were used in different studies (Chen et al., 2020; Hou et al., 2020; Jiao and Lu, 2020; Nyirabuhoro et al., 2020). For instance, some studies defined the collection of species with $>0.5\%$ relative abundance as abundant, while the ones with $<0.05\%$ relative abundance as rare (Chen et al., 2020; Hou et al., 2020), whereas in another study the species with $>0.1\%$ relative abundance were considered as abundant and the ones $<0.01\%$ as rare (Jiao and Lu, 2020). In this study, the top ranked microbial

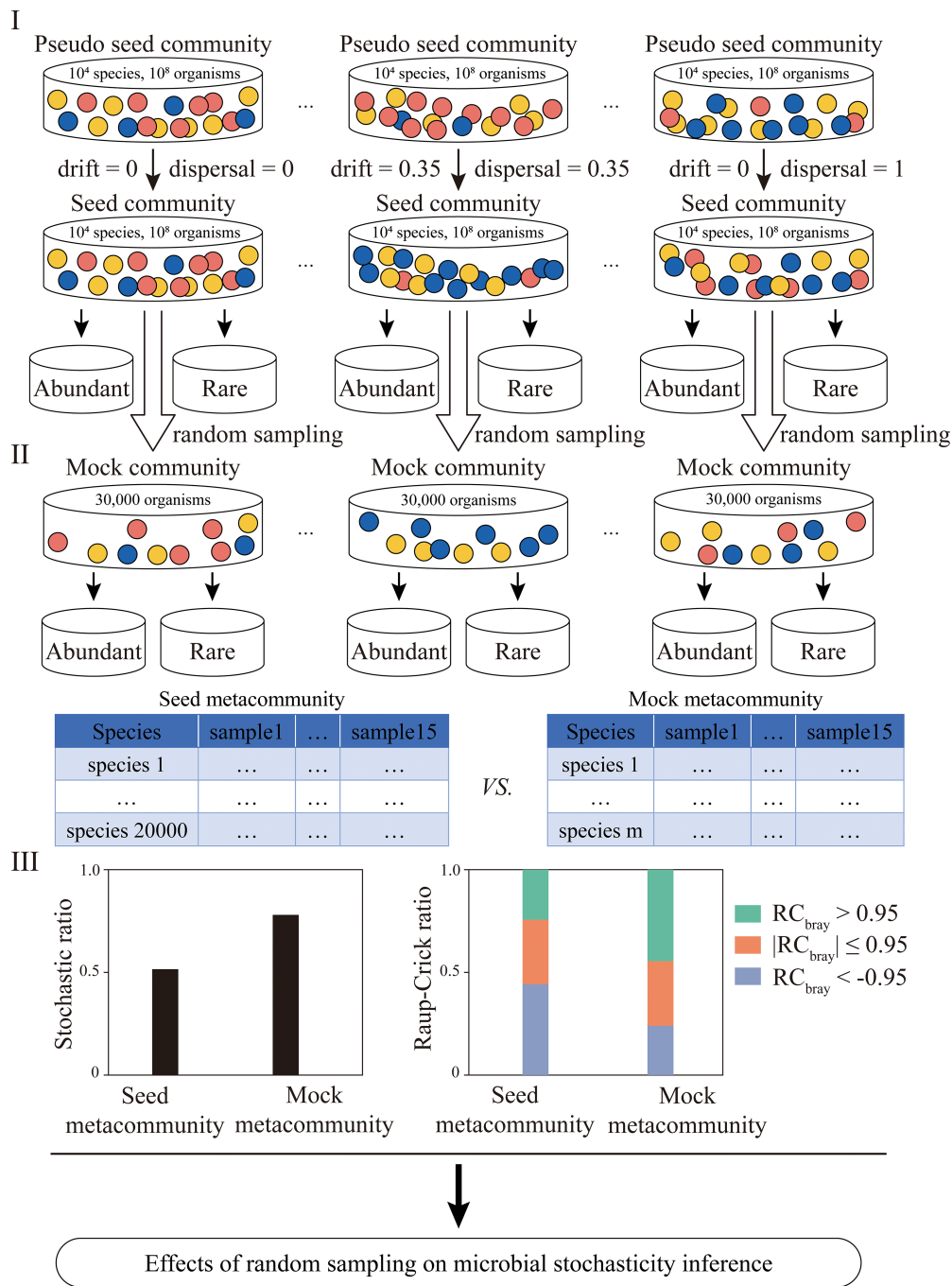


FIGURE 1

The flowchart for investigating the effects of random sampling issues on microbial stochasticity inference. First, fifteen pseudo seed communities containing 10⁴ microbial taxa and 10⁸ organisms were created. A select portion of microbial taxa were renamed and/or randomly shuffled (Supplementary Table 1), yielding seed communities with different levels of dispersal and drift. Second, mock communities with different sequencing depths were generated by randomly picking 5,000, 10,000, 30,000, 50,000, 70,000, 100,000 and 200,000 sequences from the seed communities. Third, the stochastic ratio and Raup-Crick metric were employed to assess the stochasticity of the seed metacommunity and mock metacommunity, with the difference between them representing the effect of random sampling. Microbial taxa accounting for 80% of the total relative abundance were defined as abundant subcommunity, and the rest were defined as rare subcommunity. The effect of random sampling on abundant and rare subcommunities was also investigated.

taxa who contributed to 80% total relative abundance were defined as abundant, while the rest as rare. Notably, all these criteria satisfy the basic rule of species abundance distribution in community

ecology, i.e., the vast majority abundance of microorganisms is dominated by only a few microbial species (Lynch and Neufeld, 2015). Although the abundant and rare taxa identified by different

methods may slightly differ, we did not expect strong effect of them on stochasticity analyses.

Randomization methods to generate null communities

Null models are commonly used to quantitatively disentangle the relative importance of deterministic vs. stochastic processes in structuring the compositional variations of microbial communities. Two different types of randomization methods were employed to generate null communities. The first method shuffles community composition by holding the local diversity and regional diversity close to a constant (Chase et al., 2011; Zhou et al., 2014). Here, the regional species pool is defined as the total number of microbial taxa found in all of the simulated communities with the same SeqD. Dissimilar null communities were expected. The second method draws an individual into a given taxon with a chance proportional to the relative abundance of that taxon in the regional species pool, i.e., all local communities, and at the meanwhile the local diversity and regional diversity are close to a constant (Stegen et al., 2013, 2015). As such, low compositional variations for null communities were expected. The “taxo. Null” function in the R package “NST” was used to generate different types of null communities (Ning et al., 2019). For the first randomization method, parameters including “sp. freq=prop, samp. Rich=fix, abundance=shuffle” were used. For the second randomization method, parameters including “sp. freq=prop, samp. Rich=fix, abundance=region” were used.

Microbial stochasticity inference using the stochastic ratio approach

Two different approaches were employed to evaluate the effects of random sampling issues on microbial community stochasticity inference. The first one is stochastic ratio analyses (Zhou et al., 2014; Zhou and Ning, 2017; Guo et al., 2018), which was a recently developed approach to quantitatively measure the relative importance of deterministic vs. stochastic processes in structuring the compositional variations of microbial communities. Two kinds of situations were considered in stochastic ratio calculation. First, if communities are governed by deterministic factors leading to more similar communities, the observed community similarity (C_{ij}) between the i - and j -th communities shall be greater than the null expectations (\overline{E}_{ij}). Second, if communities are governed by deterministic factors that makes communities more dissimilar, the observed community similarity (C_{ij}) between the i - and j -th communities shall be smaller than the null expectations (\overline{E}_{ij}). That being said, the observed dissimilarity ($D_{ij} = 1 - C_{ij}$) shall be greater than the null model dissimilarity ($\overline{G}_{ij} = 1 - \overline{E}_{ij}$). The stochastic ratio can therefore be calculated according to the following functions:

$$ST_{ij} = \frac{\overline{E}_{ij}}{C_{ij}}; \text{if } \overline{E}_{ij} < C_{ij}. (1)$$

$$ST_{ij} = \frac{(1 - \overline{E}_{ij})}{(1 - C_{ij})}; \text{if } \overline{E}_{ij} \geq C_{ij}. (2)$$

For each type of the abovementioned randomization methods, a total of 1,000 iterations were carried out. The null expectations were calculated by averaging similarity values across these 1,000 null communities. The modified function “tNST” in the R package “NST” to include “shuffle” option in the “abundance” parameter in the source code was used for stochastic ratio analysis (Ning et al., 2019).

Microbial stochasticity analyses using the RC_{bray} metric

In addition to the stochastic ratio approach, the RC_{bray} metric was also employed to quantify the contribution of different ecological processes to the compositional variations of microbial communities. A similar procedure as described previously was used (Chase and Myers, 2011; Stegen et al., 2013, 2015). Because it was technically almost impossible to simulate the phylogenetic relationships representing the community assembly process of mock communities, null model analysis based on the taxonomic compositional turnover was performed here. Briefly, RC_{bray} values that characterize the magnitude of deviation between the Bray–Curtis dissimilarity values of observed and null communities were calculated. RC_{bray} values larger than 0.95 suggest greater community turnover than null expectations, meaning that deterministic factors that favor different microbes account for the compositional variations. RC_{bray} values smaller than -0.95 suggest less community turnover than null expectations, meaning that deterministic factors that favor similar microbes could be the dominant process for the compositional variations. The fraction of pairwise comparisons with $|RC_{\text{bray}}| \leq 0.95$ suggests comparable community turnover between the observed and null communities, meaning that the compositional variations is a result of stochastic processes. The R function “Raup_Crick_Abundance.r” provided by Stegen et al. (2013)¹ was used for RC_{bray} metric analysis.

Results

Undersampled microbial profiles dramatically deviated from full profiles.

By comparing the compositional variations of mock communities with the seed communities, we investigated whether

¹ https://github.com/stegen/Stegen_eta_ISME_2013

and how undersampled microbial profiles deviated from full profiles. Here, 15 seed communities following lognormal distribution and with different levels of β -diversity were generated. Each seed community was composed of 10^4 species and 10^8 organisms. As a result, seed communities with β -diversity ranging from 0.07 to 0.88 were generated (Supplementary Table 1). Mock communities were then generated by random subsampling a select number of organisms from the seed communities. Here, the seed communities with 0.35 shuffling rate and 0.35 new taxa were randomly selected to illustrate the deviation of undersampled microbial profiles from full profiles (Supplementary Table 1). As a result, a large number of rare taxa (3,228~) were not captured by the mock communities, whereas the abundant taxa were rarely affected (Figures 2A–C).

The β -diversity for the seed communities and the mock communities was also comparatively analyzed. Overestimated β -diversity was observed for the undersampled mock communities (~ 0.11), including the whole community, the abundant, and rare subcommunities (Figures 2D–F). Among these, the β -diversity for rare subcommunities (~ 0.20) was the most overestimated (Figure 2F), while the β -diversity for abundant subcommunities (~ 0.08) was only slightly overestimated (Figure 2E). Notably, increasing SeqD from 50,000 to 200,000 can only slightly ease the situation of overestimated β -diversity (Figures 2D,F), suggesting that the random sampling issues associated with microbial profiling could be persistent with current and near future technologies.

The β -diversity of null mock communities was also affected

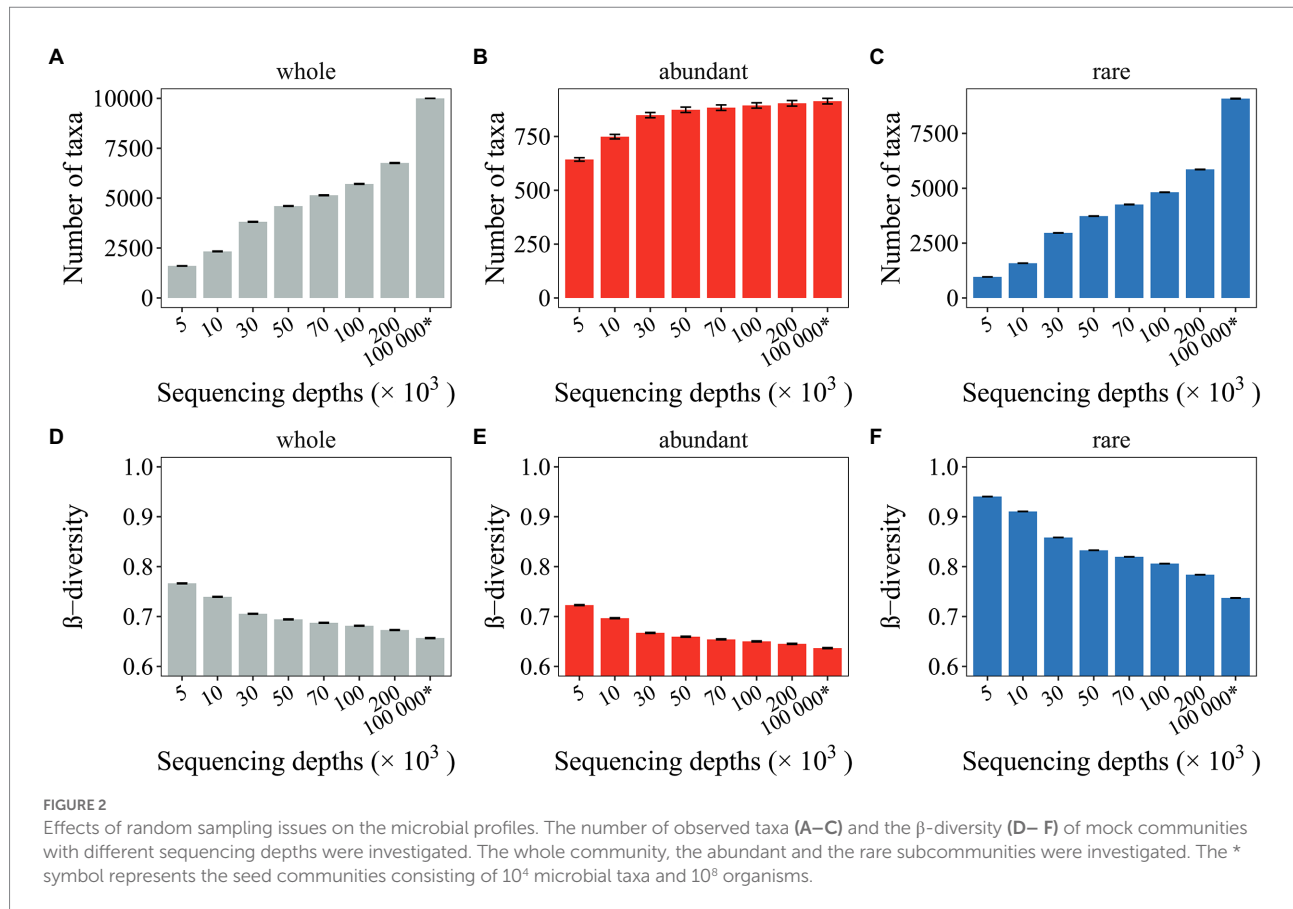
We then investigated how random sampling affected the β -diversity of null communities, based on which microbial stochasticity is inferred. Two types of commonly used randomization methods in microbial community analyses were investigated here.

As a result, deviated β -diversity of null communities was also observed. Several issues were noticed here (Figure 3). First, as expected, the β -diversity of null communities relative to observed values differed with different randomization methods. For instance, when the community composition was randomly shuffled under constraints, the β -diversity of null communities (Figure 3A) was larger than the observed β -diversity (i.e., whole and SeqD = 5,000: $0.913 > 0.766$; Figures 2D–F). However, when the community composition was generated proportionally according to the relative abundance of the taxa in the regional species pool, the β -diversity of null communities (Figure 3B) was much smaller than the observed β -diversity (i.e., whole and SeqD = 5,000: $0.724 < 0.766$; Figures 2D–F). Second, the β -diversity of null mock communities relative to that of null seed communities differed with different randomization methods. The β -diversity of null mock communities was smaller than the β -diversity of null seed communities when the

community composition was randomly shuffled under constraints ($\sim 0.927 < 0.935$; Figure 3A, “whole”). In contrast, opposite patterns were observed when the randomization of community composition was proportional to the relative abundance of microbial taxa in the regional species pool ($0.552 \sim > 0.455$; Figure 3B, “whole”). Such different patterns mainly resulted from rare subcommunities, whereas the abundant subcommunities were less affected (Figure 3). Importantly, such thoroughly differed β -diversity of null communities by different randomization methods may result in differed conclusions in microbial community stochasticity inference. Third, samples with low SeqD (e.g., 5,000 and 10,000) deviated more utterly, or even showed opposite pattern (Figure 3). The results suggested that different randomization methods exerted different effects on undersampled microbial profiles, and rare subcommunities were more strongly affected.

Microbial stochastic ratios were overestimated

Multiple community stochasticity inference approaches are available. Here, the stochastic ratio approach (Guo et al., 2018; Ning et al., 2019) was first evaluated to see how undersampled microbial profiles affected microbial community stochasticity. Overestimated stochastic ratio was observed for both randomization methods (Figure 4). Such overestimated stochastic ratio was persistently observed for rare subcommunities regardless of randomization methods (“shuffle” and SeqD = 200,000: $0.796 > 0.724$, “proportional” and SeqD = 200,000: $0.765 > 0.671$; Figures 4C,F). Comparing to what was observed for rare subcommunities, the effects of random sampling issues on stochastic ratio for abundant subcommunities differ by randomization methods (Figures 4B,E). The stochastic ratio for abundant subcommunities was rarely affected when the “shuffle” randomization method was used (Figure 4B). Most critically, undersampled microbial profiles may lead to dangerously deviated conclusions. For example, when the community composition was randomly shuffled under constraints, high stochastic ratio (> 0.75) was observed for both seed and mock metacommunities (Figures 4A–C). However, when the randomization of community composition was performed by drawing individual organisms proportional to the relative abundance of microbial taxa in the regional species pool, the stochastic ratio was low (~ 0.44) for the seed metacommunity, but high (> 0.59) for mock metacommunities, even for those with 200,000 SeqD (Figure 4D). Such issues also tended to occur with rare subcommunities (Figure 4F). Overall, the results here suggested that undersampled microbial profiles could lead to overestimated stochastic ratio inference, especially for rare subcommunities. Such overestimation may lead to carelessly different conclusions depending on which randomization methods was used.



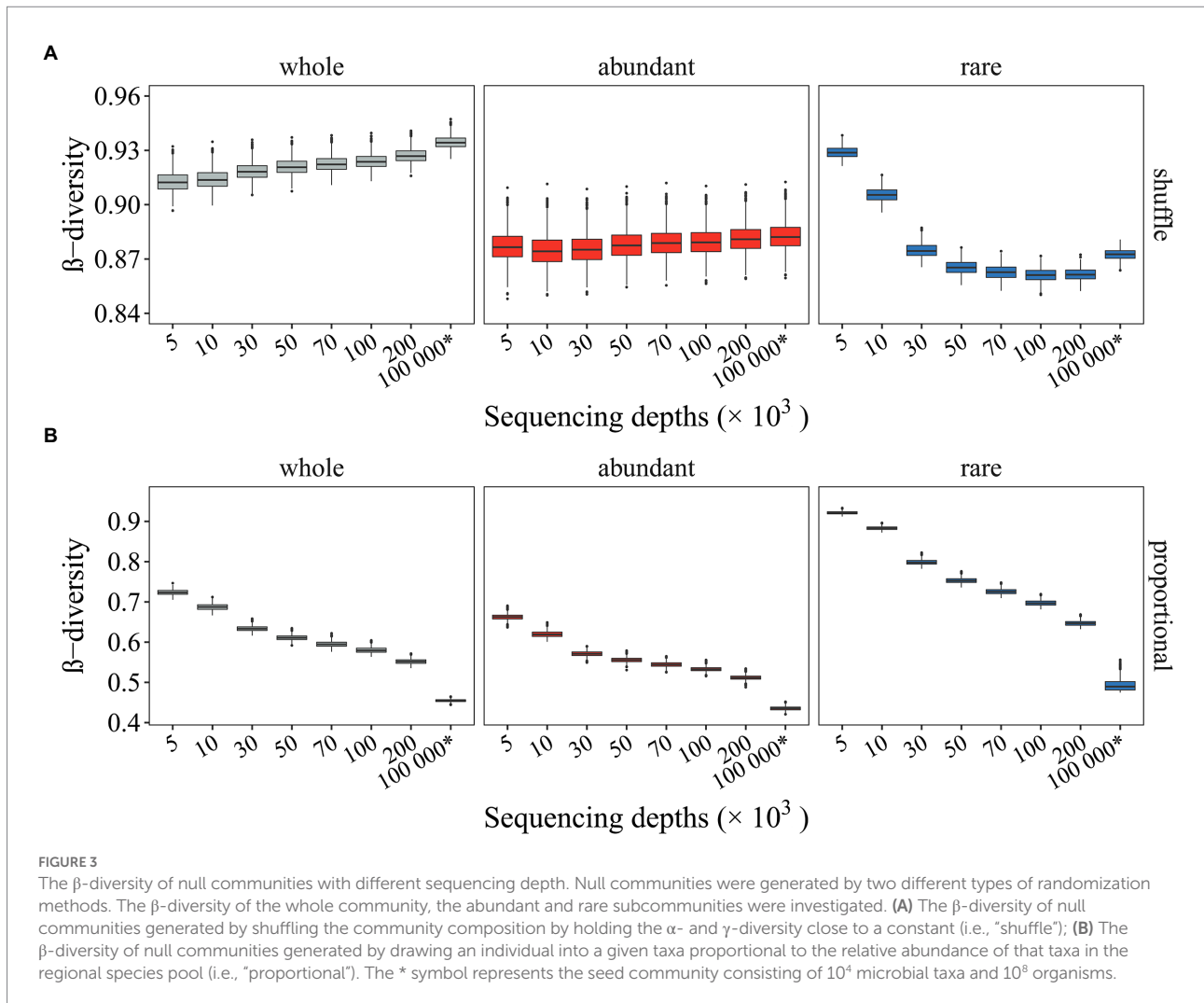
Microbial stochasticity inference using the RC_{bray} metric was also affected

In addition to stochastic ratio analyses, the RC_{bray} metric that characterizes the deviation between null distributions and observed taxonomic turnovers to infer the contributions of different processes in community assembly (Stegen et al., 2013, 2015), was also employed to evaluate how stochasticity inference was affected by random sampling issues. Notably, as it was not possible to experimentally generate the required datasets (e.g., deep sequencing of 10^8 organisms per sample), the same simulated datasets were also used here. And as it was technically almost impossible to simulate the phylogenetic relationships representing the community assembly process of mock communities, the taxonomic compositional turnover was assessed here using the RC_{bray} metric not considering the selection process inferred based on phylogenetic signals. Similarly, the same two different randomization methods (i.e., “shuffle” and “proportional”) were investigated here. Again, thoroughly different results were observed for different randomization methods (Figure 5). Such difference was mainly reflected by the relative contribution of different processes as judged by RC_{bray} values. Notably, when the “shuffle” method was used, the contribution of deterministic factors causing variable communities ($RC_{\text{bray}} > 0.95$) is overestimated, whereas the contribution of deterministic factors

causing similar communities ($RC_{\text{bray}} < -0.95$) is underestimated. Such pattern was consistently observed for the whole community, the abundant, and rare subcommunities (Figures 5A–C). However, when the “proportional” randomization method was used, overestimation of stochastic processes was observed for the rare subcommunities (Figure 5F). For the whole and abundant subcommunities, deterministic factors causing variable communities were found as the sole process responsible for the compositional variations of the mock and seed metacommunities when SeqD is larger than 70,000 (Figures 5D,E). The results suggested that RC_{bray} metric is relatively robust to random sampling issues, but could be strongly affected by randomization methods.

Discussion

Random sampling is a common issue in community ecology as complete sampling is not feasible for large-scale ecosystems or highly diverse communities. This issue becomes more critical in microbial community ecology that almost each step for profiling microbial communities is associated with random processes (Tu, 2020), resulting in undersampled microbial profiles. Previous studies suggest that such random sampling issues affect both the α - and β -diversity estimations of complex microbial communities

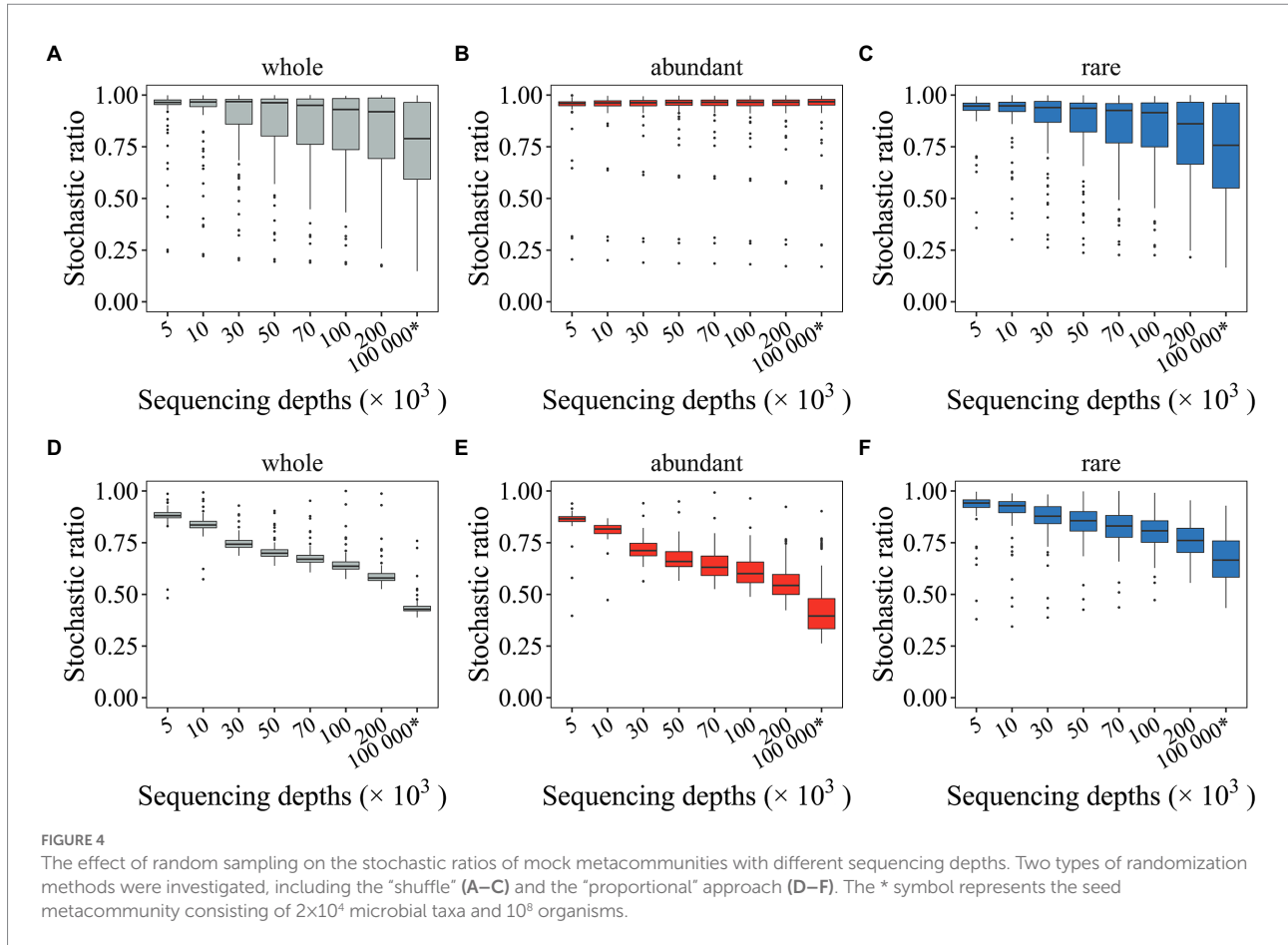


(Zhou et al., 2011, 2013; Zhan et al., 2014b). The reproducibility could be as low as 17.2% for two technical replicates and 8.2% for three technical replicates, as revealed by 16S rRNA gene amplicon sequencing using 454 pyrosequencing (Zhou et al., 2011). Our recent study suggest that random sampling issues not only affect the α - and β -diversity, but also ecological mechanisms inferred based on these indices, such as spatial scaling laws of microbial communities (Tu, 2020).

In this study, we show that microbial stochasticity inference using null model approaches is also affected by random sampling issues. The inferred community stochasticity for the whole communities, the abundant and the rare subcommunities was all affected due to random sampling issues. This was an especially critical issue for rare subcommunities, whose community stochasticity was persistently thoroughly affected regardless of which null model was used. This was in general consistence with a previous study that random sampling issues mainly affected the reproducibility of rare microbial taxa (Zhan et al., 2014a). As more studies are being made to disentangle the relative importance of deterministic vs. stochastic processes in driving the abundant and

rare subcommunities (Jiao et al., 2017; Mo et al., 2018; Xue et al., 2018; Wan et al., 2021), we urge cautions shall be made when interpreting null model results, especially for rare subcommunities.

Different randomization methods to generate null models may lead to different conclusions in microbial stochasticity analyses (Zhou and Ning, 2017). Here, the effects of random sampling issues on microbial community stochasticity inference also thoroughly differ by the randomization methods. Such difference is mainly caused by the fact that microbial stochasticity is inferred by comparing the observed community (dis)similarity with null expectations. The two randomization methods ("shuffle" and "proportional") used in this study, respectively, generated highly dissimilar and similar null model communities. This consequently led to different conclusions in stochasticity inference. In this study, we found that stochastic ratio approach was more sensitive to random sampling issues than the RC_{bray} approach that overestimated stochastic ratio was observed irrespective of which randomization method was used. In contrast, the RC_{bray} approach was more robust to random sampling issues but more strongly affected by randomization methods.

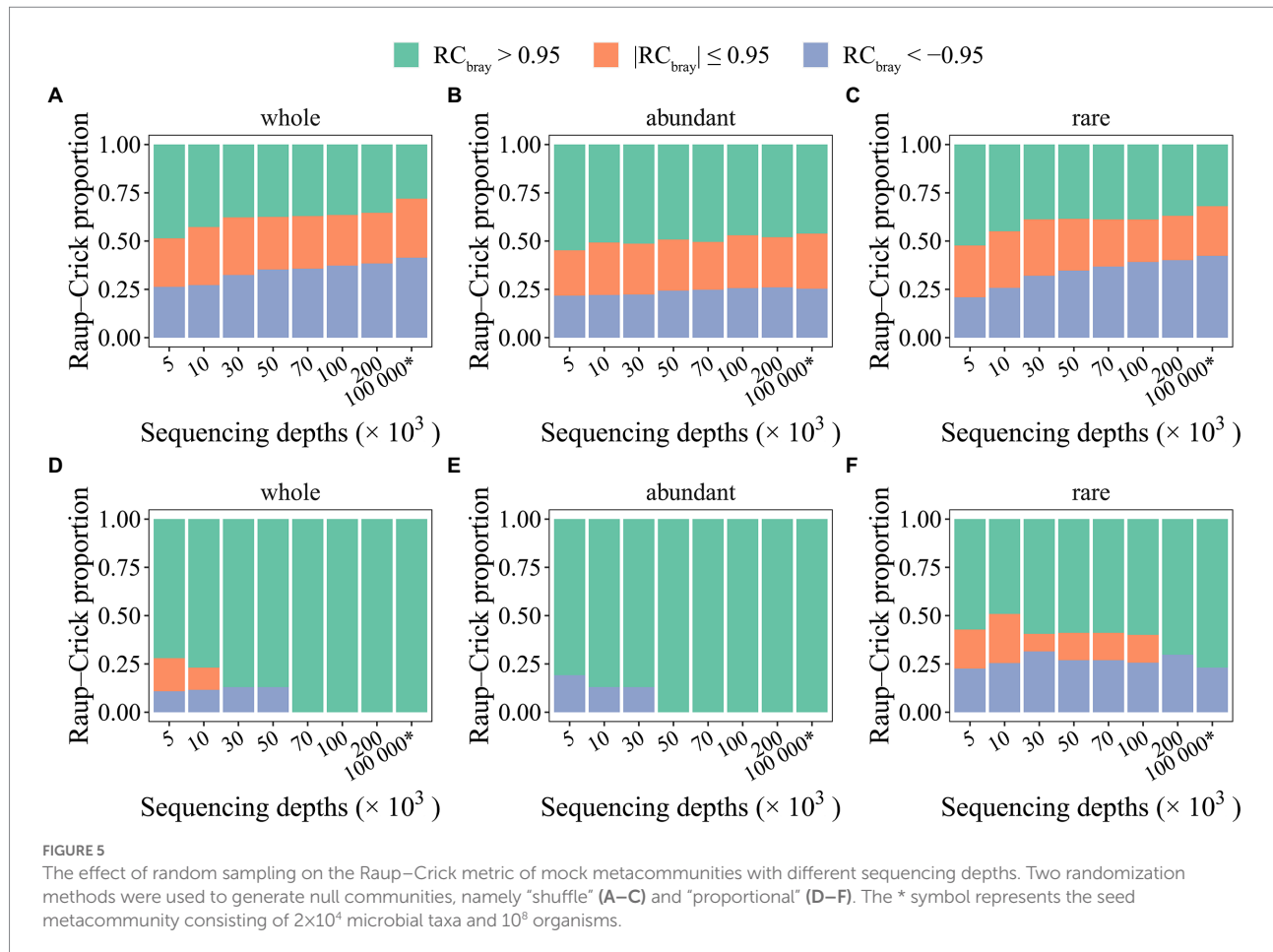


Therefore, proper selection of randomization methods for null models is also strongly recommended.

Mock (meta)communities were generated and used due to the high cost and potential technical barriers of generating the ultra-deep sequence dataset required by this study. The simulated dataset in this study is typical and representative for most environmental samples, although different microbiome types might be differently affected by random sampling issues. For instance, human microbiome that are less diverse than environmental microbiome might be less affected (Lozupone et al., 2012). The application of mock (meta)communities allows us to effectively control the variations of microbial communities and generate expected microbial profiles, simulating ecological processes such as drift and dispersal limitation (Ning et al., 2020). However, meanwhile, there are notable caveats associated with simulated datasets. First, as previously pointed out, random sampling is associated with almost all steps microbial profiles are generated, such as sample collection, DNA extraction, PCR amplification, library construction, sequencing and rarefaction (Tu, 2020). Mock (meta) communities, however, are not capable to simulate such complex procedures. In fact, generating mock (meta) communities from seed (meta)communities in the current study could be considered as a unified process anchoring the

beginning and ending status of microbial community profiling, leaving the more complex reality not thoroughly considered. Even though, strongly affected microbial stochasticity inference was observed, meaning that the real situation could be much more severe. Secondly, to our best of our knowledge, it was not possible to simulate the phylogenetic relationships representing the complex microbial community assembly processes. Therefore, the current study only considered microbial stochasticity based on taxonomic information, leaving the selection process inferred by phylogenetic signals untapped. Consequently, phylogeny-based β -diversity metrics such as UniFrac (Lozupone and Knight, 2005) and phylo-rpca (Martino et al., 2022) were also not incorporated, though it is relatively easy to incorporate different types of β -diversity metrics in null models. Nonetheless, the obtained results were still informative, showing clearly affected microbial stochasticity inference by random sampling issues associated with microbial community profiling.

Although this study focused on microbial community stochasticity, the ultimate reason causing this scenario was still the overestimated β -diversity by random sampling issues. As a result of random sampling processes associated with microbial profiling, the observed community dissimilarity (i.e., β -diversity) was overestimated, making it closer to the null community



compositions. As a result, the stochasticity for the observed communities was overestimated. Because rare subcommunities were more influenced by random sampling issues (Zhan et al., 2014a), the stochasticity of rare subcommunities was more affected than that of abundant subcommunities.

Conclusion

This study investigated the effects of random sampling issues on microbial stochasticity inference. By implementing simulated datasets, we show evidence that the stochasticity of undersampled microbial communities inferred using null models is overestimated. This issue is especially serious for rare subcommunities. Notably, such effects on the whole community and abundant communities may differ when different randomization methods are used. As more studies begin to focus on the different mechanisms governing the abundant and rare subcommunities, we urge cautions be taken when disentangling the relative importance of deterministic vs. stochastic processes, especially for rare subcommunities. Importantly, such issues could be more severe in reality, as real samples could be far more complex than simulated datasets.

Data availability statement

All custom scripts and primary data are publicly accessible on GitHub (<https://github.com/KaiMa-endeavour/Overestimated-stochasticity>).

Author contributions

QT conceptualized and designed the study. KM analyzed the data and drew the diagrams. QT and KM wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (92051110, 31971446), the National Key Research and Development Program of China (2019YFA0606700, 2020YFA0607600), the National Science Foundations of Shandong Province (ZR2020YQ21), and the Qilu Young Scholarship of Shandong University. The funders had no role in study design,

data collection and interpretation, or the decision to submit the work for publication.

Acknowledgments

The authors appreciate the editors and reviewers for their valuable comments to improve this work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Antwis, R. E., Griffiths, S. M., Harrison, X. A., Aranega-Bou, P., Arce, A., Bettridge, A. S., et al. (2017). Fifty important research questions in microbial ecology. *FEMS Microbiol. Ecol.* 93:fix 044. doi: 10.1093/femsec/fix044
- Baas-Becking, L. G. M. (1934). *Geobiologie of inleiding tot de milieukunde*. The Hague: WP Van Stockum & Zoon NV.
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U S A* 108, 4516–4522. doi: 10.1073/pnas.1000080107
- Chase, J. M. (2014). Spatial scale resolves the niche versus neutral theory debate. *J. Veg. Sci.* 25, 319–322. doi: 10.1111/jvs.12159
- Chase, J. M., Kraft, N. J. B., Smith, K. G., Vellend, M., and Inouye, B. D. (2011). Using null models to disentangle variation in community dissimilarity from variation in α -diversity. *Ecosphere* 2:art24. doi: 10.1890/es10-00117.1
- Chase, J. M., and Myers, J. A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Phil. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 366, 2351–2363. doi: 10.1098/rstb.2011.0063
- Chen, G., Wang, W., Zhang, Y., Liu, Y., Gu, X., Shi, X., et al. (2020). Abundant and rare species may invoke different assembly processes in response to climate extremes: implications for biodiversity conservation. *Ecol. Indic.* 117:106716. doi: 10.1016/j.ecolind.2020.106716
- Correa-Metrio, A., Meave, J. A., Lozano-García, S., and Bush, M. B. (2014). Environmental determinism and neutrality in vegetation at millennial time scales. *J. Veg. Sci.* 25, 627–635. doi: 10.1111/jvs.12129
- Daniel, R. (2005). The metagenomics of soil. *Nat. Rev. Microbiol.* 3, 470–478. doi: 10.1038/nrmicro1160
- Dini-Andreote, F., Stegen, J. C., van Elsland, J. D., and Salles, J. F. (2015). Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1326–E1332. doi: 10.1073/pnas.1414261112
- Dumbrell, A. J., Nelson, M., Helgason, T., Dytham, C., and Fitter, A. H. (2010). Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME J.* 4, 337–345. doi: 10.1038/ismej.2009.122
- Farjalla, V. F., Srivastava, D. S., Marino, N. A. C., Azevedo, F. D., Dib, V., Lopes, P. M., et al. (2012). Ecological determinism increases with organism size. *Ecology* 93, 1752–1759. doi: 10.1890/11-1144.1
- Fierer, N., and Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* 103, 626–631. doi: 10.1073/pnas.0507535103
- Fierer, N., Strickland, M. S., Liptzin, D., Bradford, M. A., and Cleveland, C. C. (2009). Global patterns in belowground communities. *Ecol. Lett.* 12, 1238–1249. doi: 10.1111/j.1461-0248.2009.01360.x
- Fisher, C. K., and Mehta, P. (2014). The transition between the niche and neutral regimes in ecology. *Proc. Natl. Acad. Sci. U.S.A.* 111, 13111–13116. doi: 10.1073/pnas.1405637111
- Freitas, S., Hatosy, S., Fuhrman, J. A., Huse, S. M., Mark Welch, D. B., Sogin, M. L., et al. (2012). Global distribution and diversity of marine Verrucomicrobia. *ISME J.* 6, 1499–1505. doi: 10.1038/ismej.2012.3
- Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature* 459, 193–199. doi: 10.1038/nature08058
- Galand, P. E., Casamayor, E. O., Kirchman, D. L., and Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci. U.S.A.* 106, 22427–22432. doi: 10.1073/pnas.0908284106
- Gilbert, J. A., Jansson, J. K., and Knight, R. (2014). The earth microbiome project: successes and aspirations. *BMC Biol.* 12:69. doi: 10.1186/s12915-014-0069-1
- Griffiths, R. I., Thomson, B. C., James, P., Bell, T., Bailey, M., and Whiteley, A. S. (2011). The bacterial biogeography of British soils. *Environ. Microbiol.* 13, 1642–1654. doi: 10.1111/j.1462-2920.2011.02480.x
- Guo, X., Feng, J., Shi, Z., Zhou, X., Yuan, M., Tao, X., et al. (2018). Climate warming leads to divergent succession of grassland microbial communities. *Nat. Clim. Chang.* 8, 813–818. doi: 10.1038/s41558-018-0254-2
- Harpole, W. S., and Tilman, D. (2007). Grassland species loss resulting from reduced niche dimension. *Nature* 446, 791–793. doi: 10.1038/nature05684
- Holt, R. D. (2009). Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19659–19665. doi: 10.1073/pnas.0905137106
- Hou, J., Wu, L., Liu, W., Ge, Y., Mu, T., Zhou, T., et al. (2020). Biogeography and diversity patterns of abundant and rare bacterial communities in rice paddy soils across China. *Sci. Total Environ.* 730:139116. doi:10.1016/j.scitotenv.2020.139116
- Hubbell, S. P. (1979). Tree dispersion, abundance, and diversity in a tropical dry Forest. *Science* 203, 1299–1309. doi: 10.1126/science.203.4387.1299
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, N.J.; Oxford: Princeton University Press.
- Hubbell, S. P. (2011). *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. Princeton, N.J.; Oxford: Princeton University Press.
- Jia, X., Dini-Andreote, F., and Falcão Salles, J. (2018). Community assembly processes of the microbial rare biosphere. *Trends Microbiol.* 26, 738–747. doi: 10.1016/j.tim.2018.02.011
- Jiao, S., Chen, W., and Wei, G. (2017). Biogeography and ecological diversity patterns of rare and abundant bacteria in oil-contaminated soils. *Mol. Ecol.* 26, 5305–5317. doi: 10.1111/mec.14218
- Jiao, S., and Lu, Y. (2020). Abundant fungi adapt to broader environmental gradients than rare fungi in agricultural fields. *Glob. Chang. Biol.* 26, 4506–4520. doi: 10.1111/gcb.15130
- Kylafis, G., and Loreau, M. (2011). Niche construction in the light of niche theory. *Ecol. Lett.* 14, 82–90. doi: 10.1111/j.1461-0248.2010.01551.x
- Liang, Y., Ning, D., Lu, Z., Zhang, N., Hale, L., Wu, L., et al. (2020). Century long fertilization reduces stochasticity controlling grassland microbial community succession. *Soil Biol. Biochem.* 151:108023. doi: 10.1016/j.soilbio.2020.108023
- Lozupone, C., and Knight, R. (2005). Uni Frac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005
- Lozupone, C. A., and Knight, R. (2007). Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11436–11440. doi: 10.1073/pnas.0611525104

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1011269/full#supplementary-material>

- Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature* 489, 220–230. doi: 10.1038/nature11550
- Luan, L., Jiang, Y., Cheng, M., Dini-Andreote, F., Sui, Y., Xu, Q., et al. (2020). Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. *Nat. Commun.* 11:6406. doi: 10.1038/s41467-020-20271-4
- Lynch, M. D. J., and Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* 13, 217–229. doi: 10.1038/nrmicro3400
- Lyons, K. G., Brigham, C. A., Traut, B. H., and Schwartz, M. W. (2005). Rare species and ecosystem functioning. *Conserv. Biol.* 19, 1019–1024. doi: 10.1111/j.1523-1739.2005.00106.x
- Lyons, K. G., and Schwartz, M. W. (2001). Rare species loss alters ecosystem function – invasion resistance. *Ecol. Lett.* 4, 358–365. doi: 10.1046/j.1461-0248.2001.00235.x
- Martino, C., McDonald, D., Cantrell, K., Dilmore, A. H., Vázquez-Baeza, Y., Shenav, L., et al. (2022). Compositionally aware phylogenetic Beta-diversity measures better resolve microbiomes associated with phenotype. *mSystems* 7, e0005022–e0000022. doi: 10.1128/mSystems.00050-22
- Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112. doi: 10.1038/nrmicro1341
- May, F., Gerstner, K., McGlinn, D. J., Xiao, X., and Chase, J. M. (2018). Mobsim: an R package for the simulation and measurement of biodiversity across spatial scales. *Methods Ecol. Evol.* 9, 1401–1408. doi: 10.1111/2041-210x.12986
- Mo, Y., Zhang, W., Yang, J., Lin, Y., Yu, Z., and Lin, S. (2018). Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. *ISME J.* 12, 2198–2210. doi: 10.1038/s41396-018-0153-6
- Mouillot, D., Bellwood, D. R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., et al. (2013). Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol.* 11:e1001569. doi: 10.1371/journal.pbio.1001569
- Ning, D., Deng, Y., Tiedje, J. M., and Zhou, J. (2019). A general framework for quantitatively assessing ecological stochasticity. *Proc. Natl. Acad. Sci. U.S.A.* 116, 16892–16898. doi: 10.1073/pnas.1904623116
- Ning, D., Yuan, M., Wu, L., Zhang, Y., Guo, X., Zhou, X., et al. (2020). A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* 11:4717. doi: 10.1038/s41467-020-18560-z
- Nyirabuhoro, P., Liu, M., Xiao, P., Liu, L., Yu, Z., Wang, L., et al. (2020). Seasonal variability of conditionally rare taxa in the water column Bacterioplankton Community of Subtropical Reservoirs in China. *Microb. Ecol.* 80, 14–26. doi: 10.1007/s00248-019-01458-9
- Oftner, I. D., Lunn, M., Curtis, T. P., Wells, G. F., Criddle, C. S., Francis, C. A., et al. (2010). Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15345–15350. doi: 10.1073/pnas.1000604107
- Oliverio, A., Geisen, S., Delgado-Baquerizo, M., Maestre, F., Turner, B., and Fierer, N. (2020). The global-scale distributions of soil protists and their contributions to belowground systems. *Science. Advances* 6:eaa8787. doi: 10.1126/sciadv.aax8787
- O'Malley, M. A. (2007). The nineteenth century roots of 'everything is everywhere'. *Nat. Rev. Microbiol.* 5, 647–651. doi: 10.1038/nrmicro1711
- Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., et al. (2015). Open science resources for the discovery and analysis of Tara oceans data. *Scientific Data* 2:150023. doi: 10.1038/sdata.2015.23
- Raup, D. M., and Crick, R. E. (1979). Measurement of faunal similarity in paleontology. *J. Paleontol.* 53, 1213–1227.
- Scheffer, M., and van Nes, E. H. (2006). Self-organized similarity, the evolutionary emergence of groups of similar species. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6230–6235. doi: 10.1073/pnas.0508024103
- Shen, C., Xiong, J., Zhang, H., Feng, Y., Lin, X., Li, X., et al. (2013). Soil pH drives the spatial distribution of bacterial communities along elevation on Changbai Mountain. *Soil Biol. Biochem.* 57, 204–211. doi: 10.1016/j.soilbio.2012.07.013
- Shi, Y., Li, Y., Xiang, X., Sun, R., Yang, T., He, D., et al. (2018). Spatial scale affects the relative role of stochasticity versus determinism in soil bacterial communities in wheat fields across the North China plain. *Microbiome* 6:27. doi: 10.1186/s40168-018-0409-4
- Shoemaker, W. R., Locey, K. J., and Lennon, J. T. (2017). A macroecological theory of microbial biodiversity. *Nat. Ecol. Evol.* 1:107. doi: 10.1038/s41559-017-0107
- Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U.S.A.* 103, 12115–12120. doi: 10.1073/pnas.0605127103
- Song, W., Liu, J., Qin, W., Huang, J., Yu, X., Xu, M., et al. (2022). Functional traits resolve mechanisms governing the assembly and distribution of nitrogen-cycling microbial communities in the Global Ocean. *MBio* 13:e0383221. doi: 10.1128/mbio.03832-21
- Stegen, J. C., Lin, X., Fredrickson, J. K., Chen, X., Kennedy, D. W., Murray, C. J., et al. (2013). Quantifying community assembly processes and identifying features that impose them. *ISME J.* 7, 2069–2079. doi: 10.1038/ismej.2013.93
- Stegen, J. C., Lin, X., Fredrickson, J. K., and Konopka, A. E. (2015). Estimating and mapping ecological processes influencing microbial community assembly. *Front. Microbiol.* 6, 370. doi: 10.3389/fmicb.2015.00370
- Stegen, J. C., Lin, X., Konopka, A. E., and Fredrickson, J. K. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* 6, 1653–1664. doi: 10.1038/ismej.2012.22
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348:1261359. doi: 10.1126/science.1261359
- Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi: 10.1038/nature24621
- Torsvik, V., and Øvreås, L. (2002). Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.* 5, 240–245. doi: 10.1016/s1369-5274(02)00324-7
- Tu, Q. (2020). Random sampling in metagenomic sequencing leads to overestimated spatial scaling of microbial diversity. *Environ. Microbiol.* 22, 2140–2149. doi: 10.1111/1462-2920.14973
- Tu, Q., Deng, Y., Yan, Q., Shen, L., Lin, L., He, Z., et al. (2016). Biogeographic patterns of soil diazotrophic communities across six forests in the North America. *Mol. Ecol.* 25, 2937–2948. doi: 10.1111/mec.13651
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature* 449, 804–810. doi: 10.1038/nature06244
- Vandermeer, J. H. (1972). Niche theory. *Annu. Rev. Ecol. Syst.* 3, 107–132. doi: 10.1146/annurev.ecs.03.110172.000543
- Wan, W., Gadd, G. M., Yang, Y., Yuan, W., Gu, J., Ye, L., et al. (2021). Environmental adaptation is stronger for abundant rather than rare microorganisms in wetland soils from the Qinghai-Tibet plateau. *Mol. Ecol.* 30, 2390–2403. doi: 10.1111/mec.15882
- Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6578–6583. doi: 10.1073/pnas.95.12.6578
- Wit, R., and Bouvier, T. (2006). 'Everything is everywhere, but the environment selects'; what did Baas Becking and Beijerinck really say? *Environ. Microbiol.* 8, 755–758. doi: 10.1111/j.1462-2920.2006.01017.x
- Wu, W., Lu, H.-P., Sastri, A., Yeh, Y.-C., Gong, G.-C., Chou, W.-C., et al. (2018). Contrasting the relative importance of species sorting and dispersal limitation in shaping marine bacterial versus protist communities. *ISME J.* 12, 485–494. doi: 10.1038/ismej.2017.183
- Xu, X., Wang, N., Lipson, D., Sinsabaugh, R., Schimel, J., He, L., et al. (2020). Microbial macroecology: in search of mechanisms governing microbial biogeographic patterns. *Glob. Ecol. Biogeogr.* 29, 1870–1886. doi: 10.1111/geb.13162
- Xue, Y., Chen, H., Yang, J. R., Liu, M., Huang, B., and Yang, J. (2018). Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. *ISME J.* 12, 2263–2277. doi: 10.1038/s41396-018-0159-0
- Zhan, A., He, S., Brown, E. A., Chain, F. J. J., Theriault, T. W., Abbott, C. L., et al. (2014a). Reproducibility of pyrosequencing data for biodiversity assessment in complex communities. *Methods Ecol. Evol.* 5, 881–890. doi: 10.1111/2041-210x.12230
- Zhan, A., Xiong, W., He, S., and Mac Isaac, H. J. (2014b). Influence of artifact removal on rare species recovery in natural complex communities using high-throughput sequencing. *PLoS One* 9:e96928. doi: 10.1371/journal.pone.0096928
- Zhang, H., Hou, F., Xie, W., Wang, K., Zhou, X., Zhang, D., et al. (2020a). Interaction and assembly processes of abundant and rare microbial communities during a diatom bloom process. *Environ. Microbiol.* 22, 1707–1719. doi: 10.1111/1462-2920.14820
- Zhang, X., Liu, S., Wang, J., Huang, Y., Freedman, Z., Fu, S., et al. (2020b). Local community assembly mechanisms shape soil bacterial β diversity patterns along a latitudinal gradient. *Nat. Commun.* 11:5428. doi: 10.1038/s41467-020-19228-4
- Zhang, W., Pan, Y., Yang, J., Chen, H., Holohan, B., Vaudrey, J., et al. (2018). The diversity and biogeography of abundant and rare intertidal marine microeukaryotes explained by environment and dispersal limitation. *Environ. Microbiol.* 20, 462–476. doi: 10.1111/1462-2920.13916
- Zhou, J., Deng, Y., Shen, L., Wen, C., Yan, Q., Ning, D., et al. (2016). Temperature mediates continental-scale diversity of microbes in forest soils. *Nat. Commun.* 7:12083. doi: 10.1038/ncomms12083

Zhou, J., Deng, Y., Zhang, P., Xue, K., Liang, Y., Van Nostrand, J. D., et al. (2014). Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc. Natl. Acad. Sci. U.S.A.* 111, E836–E845. doi: 10.1073/pnas.1324044111

Zhou, J., Jiang, Y.-H., Deng, Y., Shi, Z., Zhou, B. Y., Xue, K., et al. (2013). Random sampling process leads to overestimation of β -diversity of microbial communities. *MBio* 4, e00324–e00313. doi: 10.1128/mbio.00324-13

Zhou, J., and Ning, D. (2017). Stochastic community assembly: does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* 81, e00002–e00017. doi: 10.1128/MMBR.00002-17

Zhou, J., Wu, L., Deng, Y., Zhi, X., Jiang, Y.-H., Tu, Q., et al. (2011). Reproducibility and quantitation of amplicon sequencing-based detection. *ISME J.* 5, 1303–1313. doi: 10.1038/ismej.2011.11