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

Kai Ma
Shandong University https://orcid.org/0000-0001-7748-7089
Qichao Tu (tuqichao@outlook.com)
Joint Lab for Ocean Research and Education at Dalhousie University, Shandong University and Xiamen University https://orcid.org/0000-0002-3245-7545

Research Article

Keywords: random sampling, β-diversity, microbial community, stochasticity, null models, Raup-Crick metric

Posted Date: May 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1624310/v1

Abstract

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 demonstrated that microbial stochasticity inference was 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 no matter 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.

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 [1, 2], 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 [3–6]. However, the existence of highly diverse ecosystems such as rainforest, in which many organisms coexist in a same ecological niche[7, 8], challenges the throne of niche theory in community ecology. To solve such issues, Hubbell et al. proposed the neutral theory [1, 9], 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 [10–14]. 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 [15, 16].

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 [17–25]. Such efforts can date back as early as to 1930s when Baas Becking proposed the famous hypothesis “Everything is everywhere, but, the

environment selects" [26, 27]. 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 [28–35]. Recent studies also demonstrate that both deterministic and stochastic processes play critical roles in structuring the immense microbe world [36–39], and the question to resolve is which process is relatively more important [40, 41]. More recently, studies show that organism size [42–44] and spatial scale [45–47] are also critical factors influencing the relative importance of deterministic and stochastic processes in structuring microbial communities.

Microbial communities are substantially different from macrobial 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 [48]. The abundant subcommunity usually occupy < 20% of the total richness, but comprise > 80% in relative abundance [48–50]. Notably, studies suggest that the abundant taxa are usually abundant, whereas the rare taxa are persistently rare [51]. Such scenario also holds true when looking at more systematic microbial community data generated by the Earth Microbiome Project (EMP) [52], the Human Microbiome Project (HMP) [53], and the TARA Oceans Expedition [54]. Although low in relative abundance, recent studies suggest that the rare subcommunities execute nonnegligible ecosystem functions in the environment [55–57]. For such reasons, efforts have been made to disentangle the underlying ecological mechanisms structuring rare subcommunities [32, 58–60]. Although carried out in different ecosystems, these studies suggest that the abundant and rare subcommunities are structured by different mechanisms. For instance, the rare subcommunities in subtropical ecosystems are more structured by stochastic processes than abundant subcommunities [39, 60]. Similar patterns are also observed for microbial communities in oil-contaminated soils [61]. 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 [62].

Notably, the profiling of microbial communities is severely affected by random sampling issues, even using high throughput sequencing approaches [63–66]. 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. 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 [67–69], 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 sequencing depth.

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 [63–66]. And the effects of random sampling on abundant and rare subcommunities were dramatically different [70]. Since 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 sequencing depth, but could not be eliminated with current sequencing efforts. We therefore urge cautions be made for microbial stochasticity inference using null models.

Methods

Methodological framework

A framework was developed to investigate the effects of random sampling issues associated with microbial community profiling on community stochasticity inference (Fig. 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 (i.e., OTUs) and 10^8 organisms (i.e. sequences) were created. Based on the pseudo seed communities, seed communities with different levels of β-diversity were generated. Second, mock communities were generated by random subsampling select numbers of organisms from the seed communities, representing the microbial communities obtained in typical microbial ecological studies. Multiple sets of mock communities with different organism numbers were generated in order to investigate whether increasing sequencing depth would eliminate the effects of random sampling issues. Third, microbial profiles were generated for both seed and mock communities, based on which microbial community stochasticity was calculated. The community stochasticity for the seed and the mock communities were 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 [37, 40, 71] and the Raup-Crick (RC) metric [72–75], were employed here to evaluate how random sampling affected stochasticity inference.

Seed and mock community construction

Two kinds of simulated datasets were created in this study, including seed communities and mock communities. 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 [76]. Each seed community was composed by 10^4 taxa (i.e. OTUs) and 10^8 organisms (i.e. sequences), representing the basic microbial diversity in per unit environmental samples (e.g. soil) [69]. A select number (0–100%) of taxa in the seed communities were renamed as new taxa and/or randomly shuffled, mimicking community assembly processes such as drift and dispersal.

As a result, seed communities with different β-diversity were generated (Supplementary Table 1). Mock communities were then generated by random subsampling a select number (5×10^3 to 2×10^5) of organisms from the seed communities, representing microbial communities under different sequencing depth. Two major parameters associated with lognormal distribution, including “meanlog” and “sdlog”, were assessed here. The seed communities were found with “meanlog” of 6.81 ± 0.01 and “sdlog” of 2.20 ± 0.01, whereas the values for mock communities were respectively 1.02 ± 0.01 and 1.14 ± 0.01 (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) [52], the TARA Oceans expedition [54] and the Human Microbiome Project (HMP) [53]. R packages including mobsim [77] and GUniFrac [78] 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 [32, 36, 79, 80]. 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 [79, 80], whereas in another study the species with > 0.1% relative abundance were considered as abundant and the ones < 0.01% as rare [32]. In this study, the top ranked microbial 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 are dominated by only a few microbial species [48]. 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 constant [37, 72]. The second method draws an individual into a given taxa proportional to the relative abundance of that taxa in the regional species pool, i.e. all local communities, and at the meanwhile the local diversity and regional diversity were held constant [73, 74]. The "taxo.null" function in the R package NST was used to generate different types of null communities [81]. 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 [37, 40, 71], 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*-th and *j*-th communities shall be greater than the null expectations ($\bar{E}_{ij}$). Second, if communities are governed by deterministic factors that makes communities more dissimilar, the observed community similarity (C_{ij}) between the *i*-th and *j*-th communities shall be smaller than the null expectations ($\bar{E}_{ij}$). That being said, the observed dissimilarity ($D_{ij} = 1 - C_{ij}$) shall be greater than the null model dissimilarity ($G_{ij} = 1 - \bar{E}_{ij}$). The stochastic ratio can therefore be calculated according to the following functions:

$$ST_{ij} = \frac{\bar{E}_{ij}}{C_{ij}}; if \bar{E}_{ij} < C_{ij} \#(1)$$

$$ST_{ij} = \frac{\left(1 - \bar{E}_{ij}\right)}{\left(1 - C_{ij}\right)}; if \bar{E}_{ij} \geq C_{ij} \#(2)$$

For each type of the abovementioned randomization methods, a total of 1000 iterations were carried out. The null expectations were calculated by averaging similarity values across these 1000 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 [81].

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 [10, 73, 74]. 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 characterizes 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 $|RC_{bray}| < 0.95$ suggest 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” provided by Stegen et al. (https://github.com/stegen/Stegen_etal_ISME_2013) was used for RC_{bray} metric analysis [73].

Results

Undersampled microbial profiles dramatically deviated from full profiles.

By comparing the compositional variations of mock communities with the seed communities, we investigated whether and how undersampled microbial profiles deviated from full profiles. Here, fifteen seed communities following lognormal distribution and with different levels of β-diversity were generated. Each seed community was composed by 10^4 species and 10^8 organisms, representing the approximate prokaryotic diversity in one gram of soil or 0.2 liter of sea water [69]. A select number (0–100%) of organisms were renamed as new species and/or randomly shuffled, respectively simulating community assembly processes such as dispersal and drift. As a result, seed communities with β-diversity ranging from 7.48–87.83% were generated (Supplementary Table 1). Mock communities were then generated by random subsampling a select number of organisms from the seed communities. A series of mock communities with different organism numbers (5000 to 200,000) were generated, aiming to investigate the effects of increasing sequencing depth on eliminating random sampling issues. Here, the seed communities with 35% shuffling rate and 35% new taxa were selected to illustrate the deviation of undersampled microbial profiles from full profiles. As a result, a large number of rare taxa were not captured by the mock communities, whereas the abundant taxa were rarely affected (Fig. 2A, B and 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, including the whole community, the abundant and rare subcommunities (Fig. 2D, E and F). Among these, the β-diversity for rare subcommunities was the most dramatically overestimated (Fig. 2F), while the β-diversity for abundant subcommunities was only slightly overestimated (Fig. 2E). Notably, increasing sequencing depth from 50,000 to 200,000 can only slightly ease the situation of overestimated β-diversity (Fig. 2), 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. In the first randomization method, the composition of microbial communities was randomly shuffled while holding the community richness in each sample (α-diversity) and across all samples (γ-diversity) constant [37]. Here, the regional species pool is defined as the total number of microbial taxa found in all of the simulated communities with the same sequencing depth. Dissimilar null communities were expected. In the second randomization method, null microbial communities were generated by randomly drawing individuals into given taxa with the probability

proportional to the relative abundance in the regional species pool, in addition to preserving both α-diversity and γ-diversity [73]. As such, low compositional variations for null communities were expected.

As a result, deviated β-diversity of null communities was also observed. Several issues were noticed here (Fig. 3). First, as expected, the β-diversity of null communities relative to observed values dramatically differed with different randomization methods. For instance, when the community composition was randomly shuffled under constraints, the β-diversity of null communities (Fig. 3A) was larger than the observed β-diversity (Fig. 2B). 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 (Fig. 3B) was much smaller than the observed β-diversity (Fig. 2B). Second, the β-diversity of null mock communities relative to that of null seed communities dramatically 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 (Fig. 3A). 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 (Fig. 3B). Such different patterns mainly resulted from rare subcommunities, whereas the abundant subcommunities were less affected (Fig. 3). Importantly, such dramatically differed β-diversity of null communities by different randomization methods may result in dramatically differed conclusions in microbial community stochasticity inference. Third, samples with low sequencing depth (e.g. 5000 and 10000) deviated more dramatically, or even showed opposite pattern (Fig. 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 [71, 81] was first evaluated to see how undersampled microbial profiles affected microbial community stochasticity. Overestimated stochastic ratio was observed for both randomization methods (Fig. 4). Such overestimated stochastic ratio was persistently observed for rare subcommunities regardless of randomization methods (Fig. 4C and 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 (Fig. 4B and E). The stochastic ratio for abundant subcommunities was rarely affected when the “shuffle” randomization method was used (Fig. 4B). Most critically, undersampled microbial profiles may lead to dramatically 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 communities (Fig. 4A, B and 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 (~ 0.40) for the seed community, but high (> 0.52) for mock communities, even for those with 200,000 sequencing depth (Fig. 4D). Such issues also tended to occur with rare subcommunities (Fig. 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 dramatically 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 [73, 74], 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, dramatically different results were observed for different randomization methods (Fig. 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 ($RC_{bray} > 0.95$) is overestimated, whereas the contribution of deterministic factors causing similar communities ($RC_{bray} < -0.95$) is underestimated. Such pattern was consistently observed for the whole community, the abundant, and rare subcommunities (Fig. 5A, B, and C). However, when the “proportional” randomization method was used, overestimation of stochastic processes was observed for the rare subcommunities (Fig. 5F). For the whole and abundant subcommunities, deterministic factors causing variable communities was found as the sole process responsible for the compositional variations of the mock and seed communities when sequencing depth is larger than 50000 (Fig. 5D and 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 [66], 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 [63–65]. 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 [65]. 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 [66].

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 were all affected due to random sampling issues. This was an especially critical issue for rare subcommunities, whose community stochasticity was persistently dramatically affected regardless 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 [70]. As more studies are being made to disentangle the relative importance of deterministic vs. stochastic processes in driving the abundant and rare subcommunities [39, 60–62], 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 [40]. Here, the effects of random sampling issues on microbial community stochasticity inference also dramatically 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 no matter 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 communities were generated and used due to the inability to experimentally generate the ultra-deep sequence datasets required in this study. The application of mock communities allows us to effectively control the variations of microbial communities and generate expected microbial profiles [82]. However, at the 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 [66]. Mock communities, however, are not capable to simulate such complex procedures. In fact, generating mock communities from seed 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 much more severe. Secondly, to our best 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. 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. Since rare subcommunities were more influenced by random sampling issues [70], the stochasticity of rare subcommunities were more affected than that of abundant subcommunities.

Conclusions

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.

Declarations

Acknowledgements

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

This study was supported by the National Natural Science Foundation of China (92051110, 31971446), by National Key Research and Development Program of China (2019YFA0606700, 2020YFA0607600), by the Natural Science Foundations of Shandong Province (ZR2020YQ21), and by 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.

Author Contribution

QT conceptualized and designed the study. KM analyzed the data and drew the diagrams. QT and KM wrote the manuscript. All authors edited and approved the final manuscript.

Consent for Publication

All the authors give their consent to allow this manuscript to be published by the Journal of Microbial ecology.

Conflicts of Interest

The authors declare no competing interests

References

1. Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, N.J., Oxford
2. Vandermeer JH (1972) Niche Theory. Annu Rev Ecol Evol Syst 3:107–132. doi: 10.1146/annurev.es.03.110172.000543
3. Harpole WS, Tilman D (2007) Grassland species loss resulting from reduced niche dimension. Nature 446:791–793. doi: 10.1038/nature05684
4. Kylafis G, 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
5. O'Malley MA (2007) The nineteenth century roots of 'everything is everywhere'. Nat Rev Microbiol 5:647–651. doi: 10.1038/nrmicro1711
6. Holt RD (2009) Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives. Proc Natl Acad Sci USA 106:19659–19665. doi: 10.1073/pnas.0905137106
7. Hubbell SP (1979) Tree Dispersion, Abundance, and Diversity in a Tropical Dry Forest. Science 203:1299–1309. doi: 10.1126/science.203.4387.1299
8. Scheffer M, van Nes EH (2006) Self-organized similarity, the evolutionary emergence of groups of similar species. Proc Natl Acad Sci USA 103:6230–6235. doi: 10.1073/pnas.0508024103
9. Hubbell SP (2011) The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32). Princeton University Press, Princeton, N.J., Oxford
10. Chase JM, Myers JA (2011) Disentangling the importance of ecological niches from stochastic processes across scales. Philosophical Trans Royal Soc Lond Ser B Biol Sci 366:2351–2363. doi: 10.1098/rstb.2011.0063
11. Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH (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
12. Stegen JC, Lin X, Konopka AE, Fredrickson JK (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. ISME J 6:1653–1664. doi: 10.1038/ismej.2012.22
13. Fisher CK, Mehta P (2014) The transition between the niche and neutral regimes in ecology. Proc Natl Acad Sci USA 111:13111–13116. doi: 10.1073/pnas.1405637111
14. Ofiţeru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA, Sloan WT (2010) Combined niche and neutral effects in a microbial wastewater treatment community. Proc Natl Acad Sci USA 107:15345–15350. doi: 10.1073/pnas.1000604107
15. Chase JM (2014) Spatial scale resolves the niche versus neutral theory debate. J Veg Sci 25:319–322. doi: 10.1111/jvs.12159
16. Correa-Metrio A, Meave JA, Lozano-García S, Bush MB (2014) Environmental determinism and neutrality in vegetation at millennial time scales. J Veg Sci 25:627–635. doi: 10.1111/jvs.12129

17. Freitas S, Hatosy S, Fuhrman JA, Huse SM, Mark Welch DB, Sogin ML, Martiny AC (2012) Global distribution and diversity of marine Verrucomicrobia. ISME J 6:1499–1505. doi: 10.1038/ismej.2012.3
18. Fuhrman JA (2009) Microbial community structure and its functional implications. Nature 459:193–199. doi: 10.1038/nature08058
19. Fierer N, Strickland MS, Liptzin D, Bradford MA, Cleveland CC (2009) Global patterns in belowground communities. Ecol Lett 12:1238–1249. doi: 10.1111/j.1461-0248.2009.01360.x
20. Oliverio A, Geisen S, Delgado-Baquerizo M, Maestre F, Turner B, Fierer N (2020) The global-scale distributions of soil protists and their contributions to belowground systems. Sci Adv 6:eaax8787. doi: 10.1126/sciadv.aax8787
21. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R (2017) A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457–463. doi: 10.1038/nature24621
22. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Øvreås L, Reysenbach A-L, Smith VH, Staley JT (2006) Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 4:102–112. doi: 10.1038/nrmicro1341
23. Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. Proc Natl Acad Sci USA 104:11436–11440. doi: 10.1073/pnas.0611525104
24. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA 108:4516–4522. doi: 10.1073/pnas.1000080107
25. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci USA 103:626–631. doi: 10.1073/pnas.0507535103
26. Baas-Becking LGM (1934) Geobiologie of inleiding tot de milieukunde. WP Van Stockum & Zoon NV
27. Wit R, 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
28. Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley AS (2011) The bacterial biogeography of British soils. Environ Microbiol 13:1642–1654. doi: 10.1111/j.1462-2920.2011.02480.x
29. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, de Vargas C, Gorsky

C, Grimsley G, Hingamp N, Iudicone P, Jaillon D, Not O, Ogata F, Pesant H, Speich S, Stemmann S, Sullivan L, Weissenbach MB (2015) J Structure and function of the global ocean microbiome. Science 348: 1261359. doi: 10.1126/science.1261359

30. Tu Q, Deng Y, Yan Q, Shen L, Lin L, He Z, Wu L, Van Nostrand JD, Buzzard V, Michaletz ST, Enquist BJ, Weiser MD, Kaspari M, Waide RB, Brown JH, Zhou J (2016) Biogeographic patterns of soil diazotrophic communities across six forests in the North America. Mol Ecol 25:2937–2948. doi: 10.1111/mec.13651
31. Zhou J, Deng Y, Shen L, Wen C, Yan Q, Ning D, Qin Y, Xue K, Wu L, He Z, Voordeckers JW, Nostrand JDV, Buzzard V, Michaletz ST, Enquist BJ, Weiser MD, Kaspari M, Waide R, Yang Y, Brown JH (2016) Temperature mediates continental-scale diversity of microbes in forest soils. Nat Commun 7:12083. doi: 10.1038/ncomms12083
32. Jiao S, Lu Y (2020) Abundant fungi adapt to broader environmental gradients than rare fungi in agricultural fields. Glob Change Biol 26:4506–4520. doi: 10.1111/gcb.15130
33. Shen C, Xiong J, Zhang H, Feng Y, Lin X, Li X, Liang W, Chu H (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
34. Liang Y, Ning D, Lu Z, Zhang N, Hale L, Wu L, Clark IM, McGrath SP, Storkey J, Hirsch PR, Sun B, Zhou J (2020) Century long fertilization reduces stochasticity controlling grassland microbial community succession. Soil Biol Biochem 151:108023. doi: 10.1016/j.soilbio.2020.108023
35. Xu X, Wang N, Lipson D, Sinsabaugh R, Schimel J, He L, Soudzilovskaia NA, Tedersoo L (2020) Microbial macroecology: In search of mechanisms governing microbial biogeographic patterns. Glob Ecol Biogeogr 29:1870–1886. doi: 10.1111/geb.13162
36. Nyirabuhoro P, Liu M, Xiao P, Liu L, Yu Z, Wang L, Yang J (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
37. Zhou J, Deng Y, Zhang P, Xue K, Liang Y, Van Nostrand JD, Yang Y, He Z, Wu L, Stahl DA, Hazen TC, Tiedje JM, Arkin AP (2014) Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. Proc Natl Acad Sci USA 111:E836. doi: 10.1073/pnas.1324044111
38. Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF (2015) Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. Proc Natl Acad Sci USA 112:E1326. doi: 10.1073/pnas.1414261112
39. Xue Y, Chen H, Yang JR, Liu M, Huang B, 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
40. Zhou J, Ning D (2017) Stochastic Community Assembly: Does It Matter in Microbial Ecology? Microbiol Mol Biol Rev 81:e00002–00017. doi: 10.1128/MMBR.00002-17
41. Antwis RE, Griffiths SM, Harrison XA, Aranega-Bou P, Arce A, Bettridge AS, Brailsford FL, de Menezes A, Devaynes A, Forbes KM, Fry EL, Goodhead I, Haskell E, Heys C, James C, Johnston SR, Lewis GR,

Lewis Z, Macey MC, McCarthy A, McDonald JE, Mejia-Florez NL, O'Brien D, Orland C, Pautasso M, Reid WDK, Robinson HA, Wilson K, Sutherland WJ (2017) Fifty important research questions in microbial ecology. FEMS Microbiol Ecol 93:fix044. doi: 10.1093/femsec/fix044

42. Wu W, Lu H-P, Sastri A, Yeh Y-C, Gong G-C, Chou W-C, Hsieh C-H (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
43. Farjalla VF, Srivastava DS, Marino NAC, Azevedo FD, Dib V, Lopes PM, Rosado AS, Bozelli RL, Esteves FA (2012) Ecological determinism increases with organism size. Ecology 93:1752–1759. doi: 10.1890/11-1144.1
44. Luan L, Jiang Y, Cheng M, Dini-Andreote F, Sui Y, Xu Q, Geisen S, Sun B (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
45. Zhang X, Liu S, Wang J, Huang Y, Freedman Z, Fu S, Liu K, Wang H, Li X, Yao M, Liu X, Schuler J (2020) 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
46. Shi Y, Li Y, Xiang X, Sun R, Yang T, He D, Zhang K, Ni Y, Zhu Y-G, Adams JM, Chu H (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
47. Song W, Liu J, Qin W, Huang J, Yu X, Xu M, Stahl D, Jiao N, Zhou J, Tu Q (2022) Functional Traits Resolve Mechanisms Governing the Assembly and Distribution of Nitrogen-Cycling Microbial Communities in the Global Ocean. mBio 0:e03832–e03821. doi: 10.1128/mbio.03832-21
48. Lynch MDJ, Neufeld JD (2015) Ecology and exploration of the rare biosphere. Nat Rev Microbiol 13:217–229. doi: 10.1038/nrmicro3400
49. Zhang H, Hou F, Xie W, Wang K, Zhou X, Zhang D, Zhu X (2020) 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
50. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci USA 103:12115–12120. doi: 10.1073/pnas.0605127103
51. Galand PE, Casamayor EO, Kirchman DL, Lovejoy C (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. Proc Natl Acad Sci USA 106:22427. doi: 10.1073/pnas.0908284106
52. Gilbert JA, Jansson JK, Knight R (2014) The Earth Microbiome project: successes and aspirations. BMC Biol 12:69. doi: 10.1186/s12915-014-0069-1
53. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007) The human microbiome project. Nature 449:804–810. doi: 10.1038/nature06244
54. Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, Iudicone D, Karsenti E, Speich S, Troublé R, Dimier C, Searson S (2015) Open science resources for the discovery and analysis of Tara Oceans data. Sci data 2:150023. doi: 10.1038/sdata.2015.23

55. Lyons KG, Brigham CA, Traut BH, Schwartz MW (2005) Rare Species and Ecosystem Functioning. Conserv Biol 19:1019–1024. doi: 10.1111/j.1523-1739.2005.00106.x
56. Lyons KG, Schwartz MW (2001) Rare species loss alters ecosystem function – invasion resistance. Ecol Lett 4:358–365. doi: 10.1046/j.1461-0248.2001.00235.x
57. Mouillot D, Bellwood DR, Baraloto C, Chave J, Galzin R, Harmelin-Vivien M, Kulbicki M, Lavergne S, Lavorel S, Mouquet N, Paine CET, Renaud J, Thuiller W (2013) Rare Species Support Vulnerable Functions in High-Diversity Ecosystems. PLoS Biol 11:e1001569. doi: 10.1371/journal.pbio.1001569
58. Jia X, Dini-Andreote F, 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
59. Zhang W, Pan Y, Yang J, Chen H, Holohan B, Vaudrey J, Lin S, McManus GB (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
60. Mo Y, Zhang W, Yang J, Lin Y, Yu Z, 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
61. Jiao S, Chen W, 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
62. Wan W, Gadd GM, Yang Y, Yuan W, Gu J, Ye L, Liu W (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
63. Zhan A, Xiong W, He S, MacIsaac HJ (2014) 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
64. Zhou J, Jiang Y-H, Deng Y, Shi Z, Zhou BY, Xue K, Wu L, He Z, Yang Y (2013) Random sampling process leads to overestimation of β-diversity of microbial communities. mBio 4:e00324–e00313. doi: 10.1128/mbio.00324-13
65. Zhou J, Wu L, Deng Y, Zhi X, Jiang Y-H, Tu Q, Xie J, van Nostrand JD, He Z, Yang Y (2011) Reproducibility and quantitation of amplicon sequencing-based detection. ISME J 5:1303–1313. doi: 10.1038/ismej.2011.11
66. 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
67. Daniel R (2005) The metagenomics of soil. Nat Rev Microbiol 3:470–478. doi: 10.1038/nrmicro1160
68. Torsvik V, 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
69. Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: the unseen majority. Proc Natl Acad Sci USA 95:6578–6583. doi: 10.1073/pnas.95.12.6578

70. Zhan A, He S, Brown EA, Chain FJJ, Therriault TW, Abbott CL, Heath DD, Cristescu ME, MacIsaac HJ (2014) Reproducibility of pyrosequencing data for biodiversity assessment in complex communities. Methods Ecol Evol 5:881–890. doi: 10.1111/2041-210x.12230
71. Guo X, Feng J, Shi Z, Zhou X, Yuan M, Tao X, Hale L, Yuan T, Wang J, Qin Y, Zhou A, Fu Y, Wu L, He Z, Van Nostrand JD, Ning D, Liu X, Luo Y, Tiedje JM, Yang Y, Zhou J (2018) Climate warming leads to divergent succession of grassland microbial communities. Nat Clim Change 8:813–818. doi: 10.1038/s41558-018-0254-2
72. Chase JM, Kraft NJB, Smith KG, Vellend M, Inouye BD (2011) Using null models to disentangle variation in community dissimilarity from variation in α-diversity. Ecosphere 2:art24. doi: 10.1890/es10-00117.1
73. Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, Rockhold ML, Konopka A (2013) Quantifying community assembly processes and identifying features that impose them. ISME J 7:2069–2079. doi: 10.1038/ismej.2013.93
74. Stegen JC, Lin X, Fredrickson JK, Konopka AE (2015) Estimating and mapping ecological processes influencing microbial community assembly. Front Microbiol 6. doi: 10.3389/fmicb.2015.00370
75. Raup DM, Crick RE (1979) Measurement of Faunal Similarity in Paleontology. J Paleontol 53:1213–1227
76. Shoemaker WR, Locey KJ, Lennon JT (2017) A macroecological theory of microbial biodiversity. Nat Ecol Evol 1:107. doi: 10.1038/s41559-017-0107
77. May F, Gerstner K, McGlinn DJ, Xiao X, Chase JM (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
78. Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, Collman RG, Bushman FD, Li H (2012) Associating microbiome composition with environmental covariates using generalized UniFrac distances. Bioinformatics 28:2106–2113. doi: 10.1093/bioinformatics/bts342
79. Chen G, Wang W, Zhang Y, Liu Y, Gu X, Shi X, Wang M (2020) Abundant and rare species may invoke different assembly processes in response to climate extremes: Implications for biodiversity conservation. Ecol Ind 117:106716. doi: 10.1016/j.ecolind.2020.106716
80. Hou J, Wu L, Liu W, Ge Y, Mu T, Zhou T, Li Z, Zhou J, Sun X, Luo Y, Christie P (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
81. Ning D, Deng Y, Tiedje JM, Zhou J (2019) A general framework for quantitatively assessing ecological stochasticity. Proc Natl Acad Sci USA 116:16892–16898. doi: 10.1073/pnas.1904623116
82. Ning D, Yuan M, Wu L, Zhang Y, Guo X, Zhou X, Yang Y, Arkin AP, Firestone MK, Zhou J (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

Figures

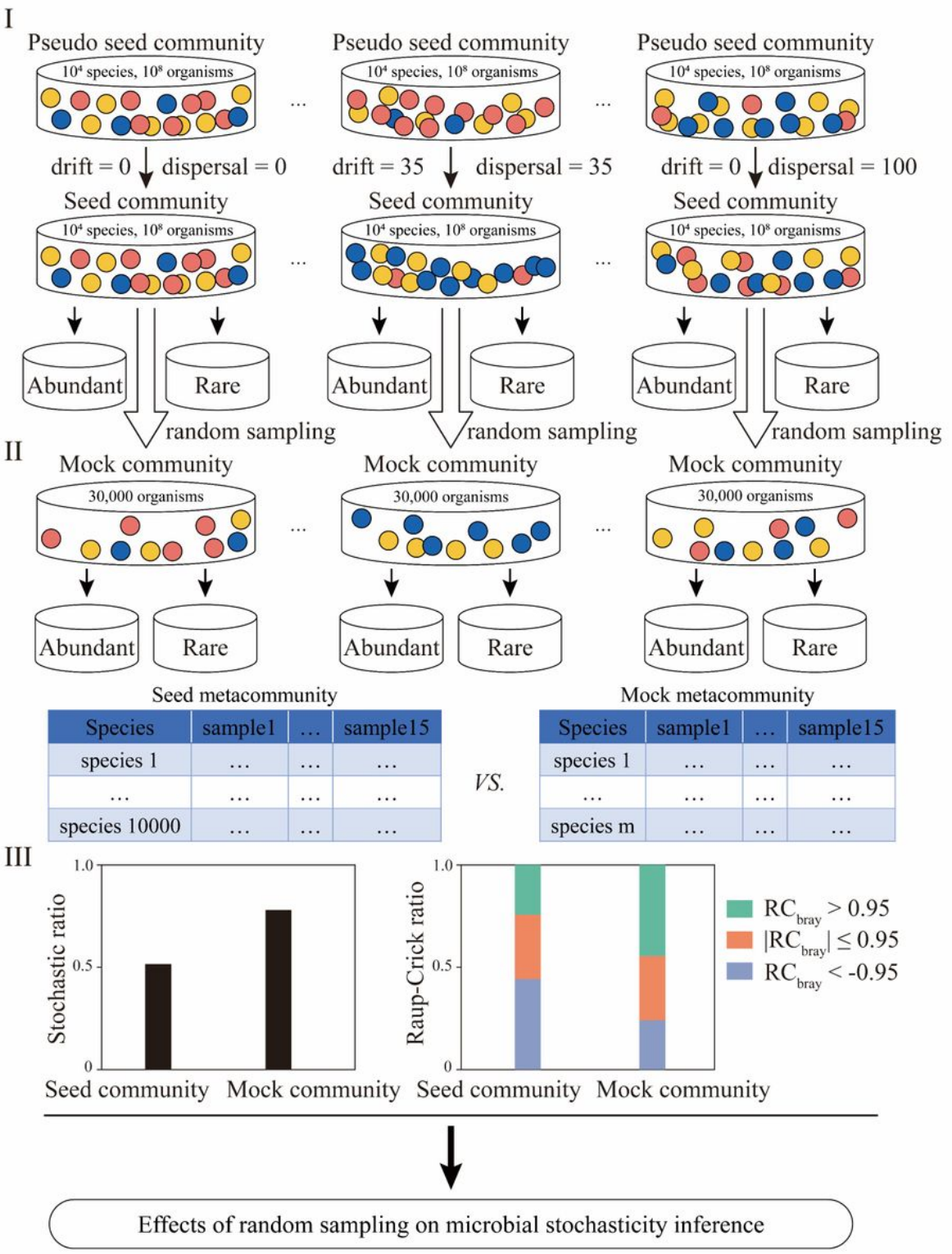

Figure 1

The flowchart for investigating the effects of random sampling issues on microbial stochasticity inference. First, fifteen pseudo seed communities containing 10^4 microbial taxa (i.e., OTUs) and 10^8 organisms (i.e. sequences) 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 communities and mock communities, 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.

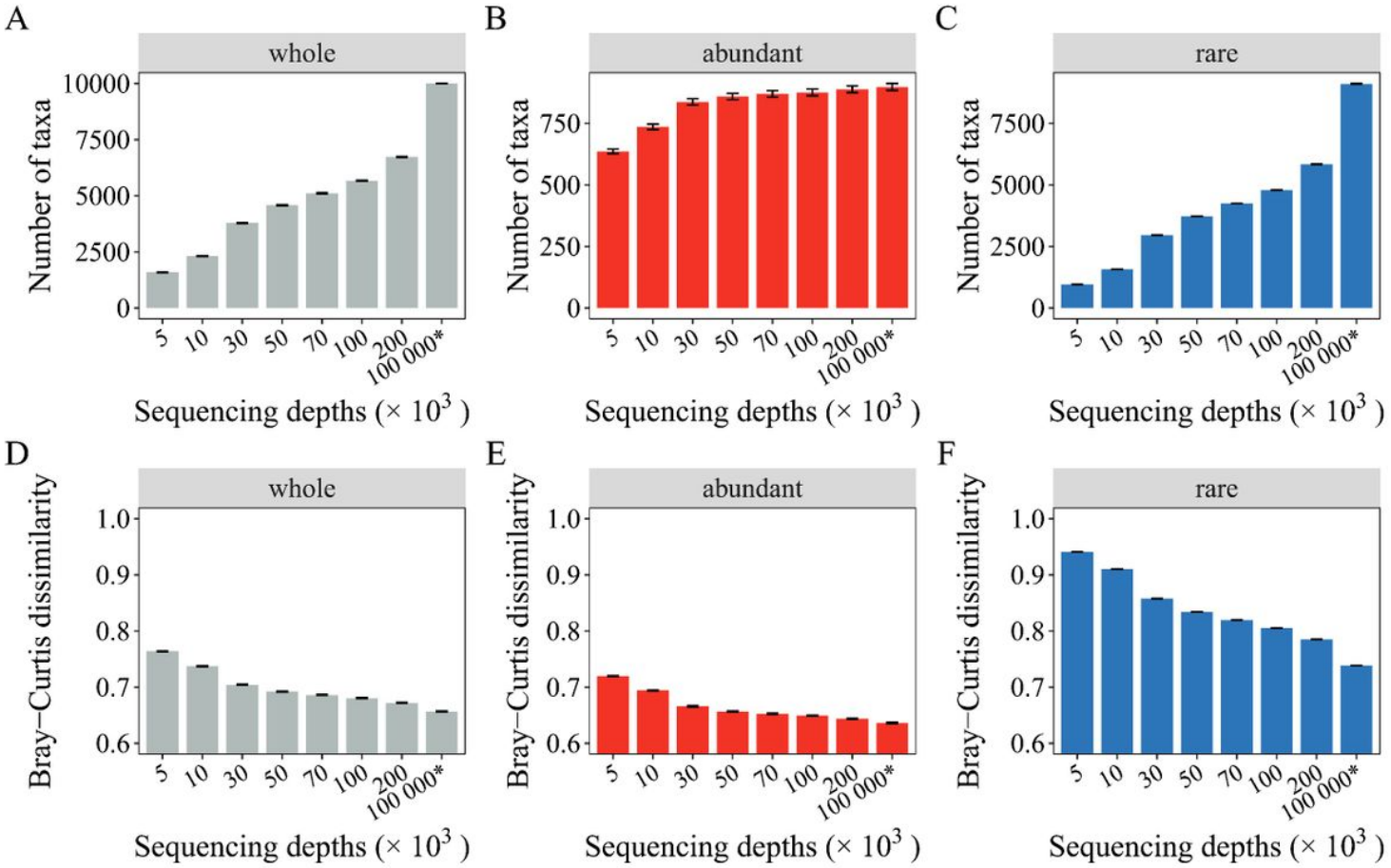

Figure 2

Effects of random sampling issues on the microbial profiles. The number of observed taxa (A, B, C) and the b-diversity (D, E, F) of mock communities with different sequencing depths were investigated. The whole community, the abundant and the rare subcommunities were investigated. The * symbol represents the seed communities consisting of 10^4 microbial taxa and 10^8 organisms.

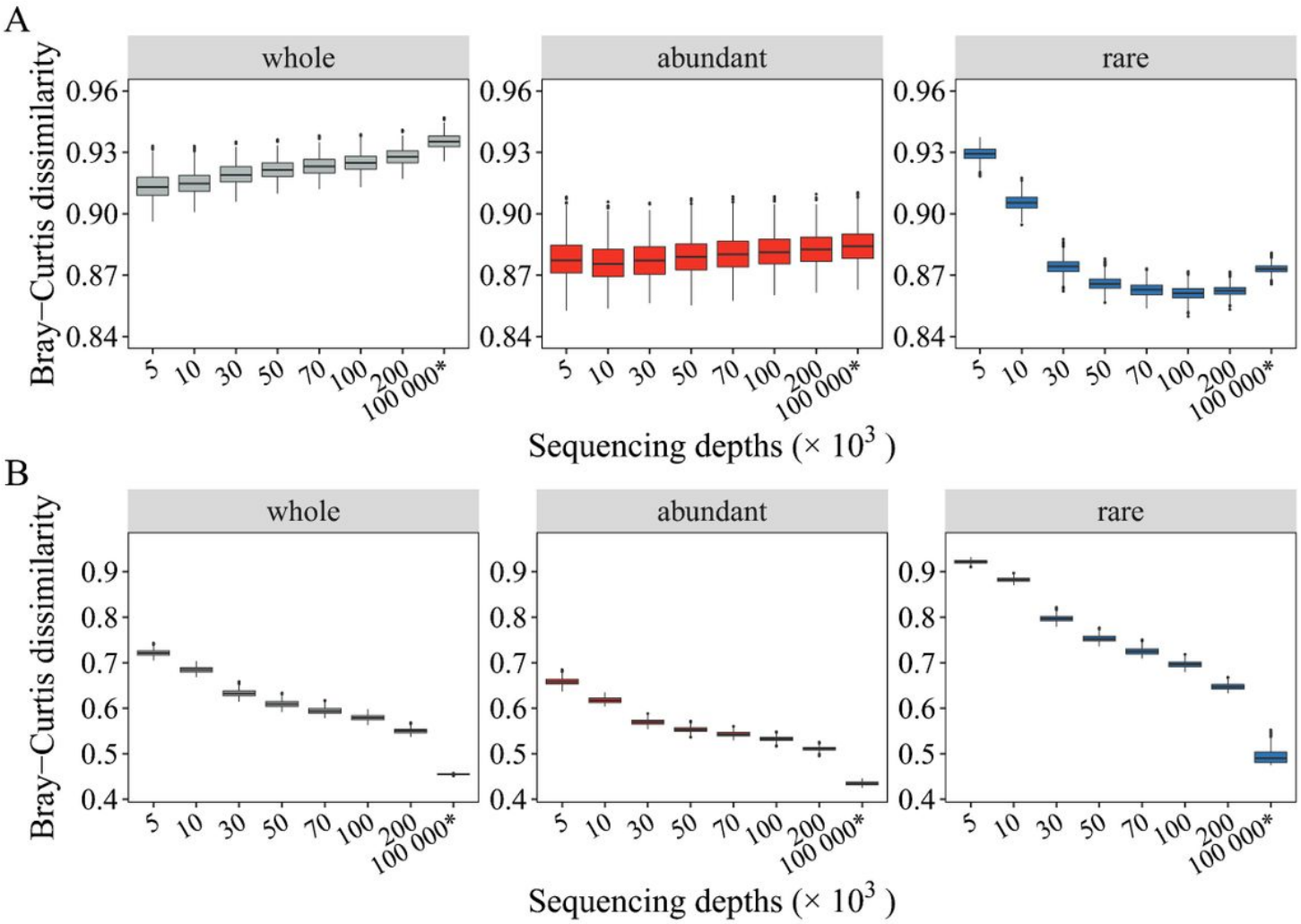

Figure 3

The b-diversity of null communities with different sequencing depth. Null communities were generated by two different types of randomization methods. The β-diversity of the whole community, the abundant and rare subcommunities were investigated. (A) The β-diversity of null communities generated by shuffling the community composition by holding the α- and γ-diversity constant (i.e. “shuffle”); (B) The β-diversity of null communities generated by drawing an individual into a given taxa proportional to the relative abundance of that taxa in the regional species pool (i.e. “proportional”). The * symbol represents the seed community consisting of 10^4 microbial taxa and 10^8 organisms.

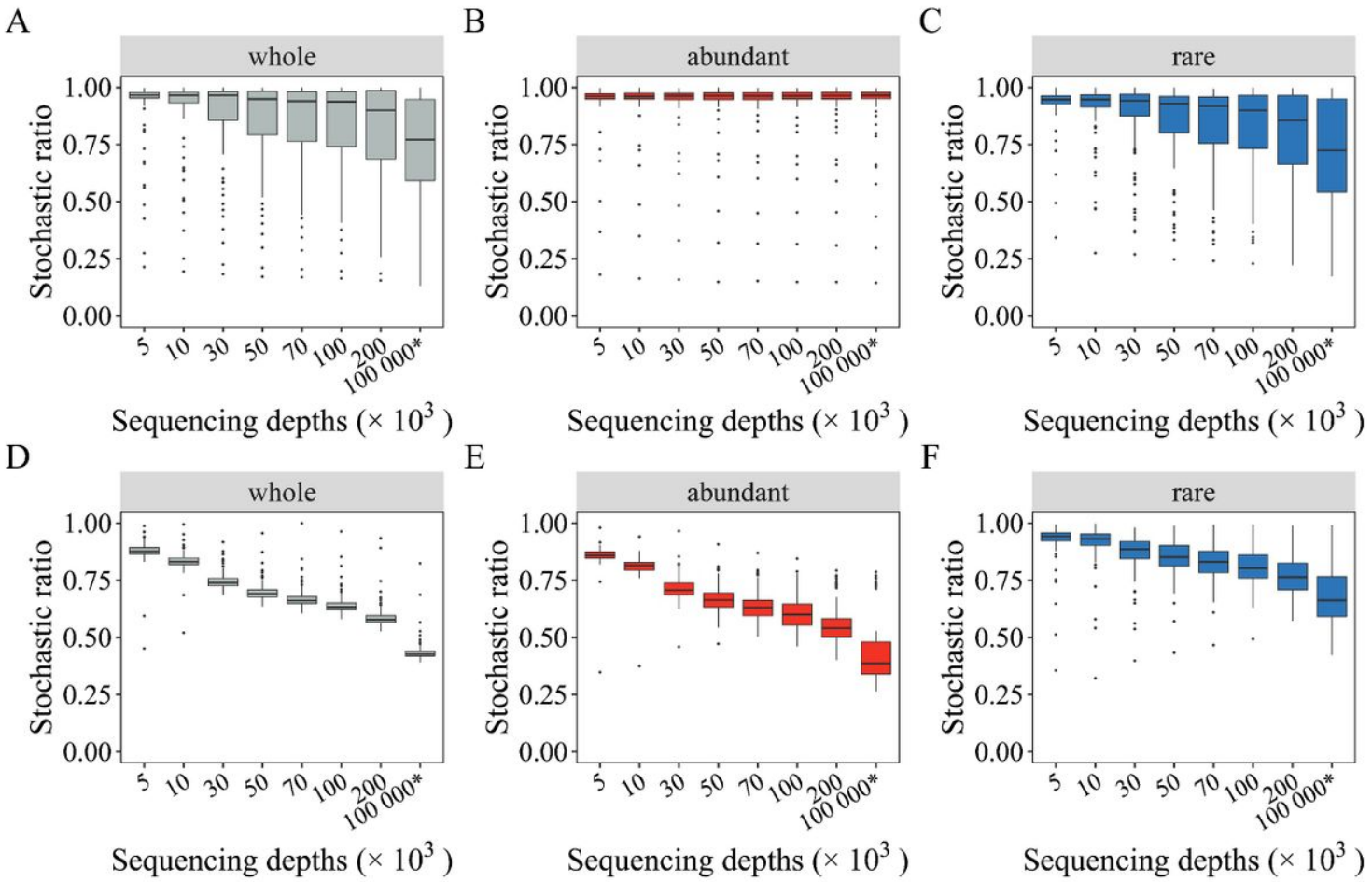

Figure 4

The effect of random sampling on the stochastic ratios of mock communities with different sequencing depths. Two types of randomization methods were investigated, including the “shuffle” (A, B, C) and the “proportional” approach (D, E, F). The * symbol represents the seed community consisting of 10^4 microbial taxa and 10^8 organisms.

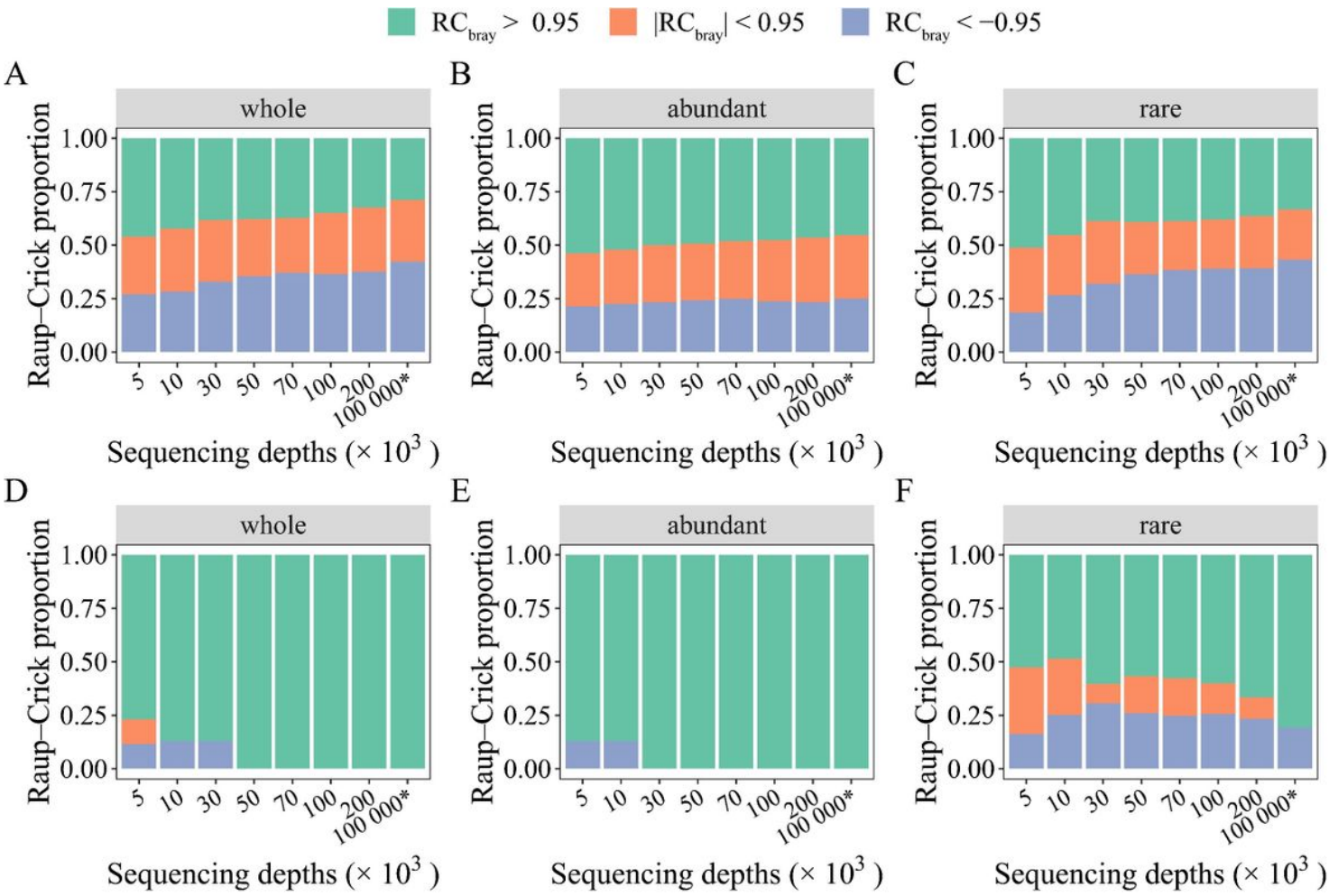

Figure 5

The effect of random sampling on the Raup-Crick metric of mock communities with different sequencing depths. Two randomization methods were used to generate null communities, namely “shuffle” (A, B and C) and “proportional” (D, E and F). The * symbol represents the seed community consisting of 10^4 microbial taxa and 10^8 organisms.

Supplementary Files

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

- supplementary.docx