

**Author comments are in bold and start with \*\*\*.**

Referee #1 (Comments to the Author):

I have carefully re-read the revised manuscript, and I can see and appreciate all the hard work that went into the revision. Unfortunately, many of my original concerns remain. However, before I get into it, I want to emphasize that I think the authors are exploring a very interesting and important topic. It is just that the topic is also very difficult to tackle, and the data/patterns need careful consideration before jumping to conclusions.

This study aims to understand the processes underlying recruitment patterns into host-associated microbial communities. The central finding is that the order of arrival (i.e., recruitment/detection) is non-random. Specifically, that species which are more closely related to community members recruit sooner than species which are less closely related to community members. While I am pleased to see many improvements to the manuscript, I still have significant lingering concerns.

My foremost concern, as I stated in my initial review, is that, “I don’t think it is acceptable nowadays to use patterns of phylogenetic relatedness in communities to infer community assembly processes in a dynamic system.” There is now strong theoretical evidence that species in a community may be underdispersed (clustered) due to either competition or environmental filtering. The link between community patterns and process depends critically on the traits and nature of competition/environmental filtering occurring in the community (e.g., Mayfield and Levine 2010).

**\*\*\* We understand the reviewer’s concern that if underdispersion can arise from either environmental filtering or from competition, it is not possible to separate the two using community phylogenetics. Per the reviewer’s advice, we take the position of Mayfield and Levine (2010), who argue that a) “over-dispersion can still only result from competitive exclusion” of close relatives, and b) environmental filtering or competitive exclusion between distant relatives can lead to underdispersion. In our paper, we reject the hypothesis that there would be competitive exclusion of close relatives in microbiome communities, insofar as a direct consequence of “competitive exclusion of close relatives” is overdispersed recruitment (as suggested above). We agree with the reviewer that rejection of that hypothesis (and of the null) leaves us with a subset of possible processes behind the pattern, and we acknowledge which processes these may be (chiefly, environmental filtering, and competition between distantly related species). Per the reviewer’s comment, we have changed our Abstract and Introduction to make this important epistemological point more clear.**

Moreover, in this paper, we are two steps further removed from process than in many other studies on community patterns: (1) we are ignoring the importance of species abundances to community and immigration dynamics,

**\*\*\*We have addressed this issue by adding the “individual-null” model the reviewer requested in their second comment and found that skin and gut microbiome datasets we analyzed are**

underdispersed regardless of whether or not we test against a null hypothesis represented by either abundance-weighted or unweighted neutral models (called “individual-null” and “species-null” in this document). We’ve included this model and result for those readers similarly curious about the species- vs. individual null models of recruitment (in the case where the abundances we observe reflect species pool abundances or propagule pressures), and included code to use the “individual null” in our tutorial vignette.

and (2) we are ignoring the fact that phylogenetic relatedness does not always correlate with functional similarity (Cavendar-Bares et al 2014).

**\*\*\*We searched for Cavendar-Bares et al. 2014, which we couldn’t find. Instead we assume that this was a minor typo, and the reviewer is referring to their 2004 paper, which found a lack of correlation between phylogenetic relatedness and functional similarity. However, that study was within a single genus of trees (*Quercus*), and for bacteria, especially those found in the human microbiome, function has been shown to correlate with 16S bacterial phylogeny (Langille et al. 2013, Zaneveld et al. 2010). This has also been shown to be very common by several studies cited in Mayfield and Levine (2010) as well as Wiens (2010).**

In other words, I am not convinced that there is anything we can conclude about process here (or at least I haven’t been sufficiently convinced of it by the Introduction). As such, I do not agree with the overall framing of the study in the Abstract and Introduction; in particular, I do not think it is worthwhile to bring up overdispersion / underdispersion / naturalization / preadaptation hypotheses if they are not readily useful for connecting pattern and process. Now, it may be that reporting the pattern of phylogenetically underdispersed recruitment is useful from an applied standpoint (and the authors raise this point in the Discussion). However, in my opinion, the Abstract/Introduction should focus on that applied angle, and not on trying to provide support for conceptual frameworks that are ultimately unable to unambiguously connect pattern and process.

**\*\*\* As stated above, we understand where the reviewer is coming from. We have edited the abstract to focus on the empirical results, i.e. that species are more likely to be recruited into human gut and skin microbiomes if a close relative is already present, and the applied aspects of our work. We have also edited our Introduction similarly. We do feel that it is necessary to explain the theoretical / historical / mechanistic reasoning that led to the hypotheses we test in our paper, but we have changed our Introduction with the reviewer’s comment in mind, especially in regard to making sure that our overdispersion / underdispersion / neutral hypotheses are clearly in regard to empirical patterns in data.**

A closely related concern involves the assumptions that go into the neutral/null model, which interfere with my ability to fully believe the overall conclusions. First: the neutral/null model in this study assumes that all species in the regional species pool are equally likely to arrive. The authors then treat any deviations from this null model as a non-random pattern of arrival. However, in reality, a true neutral/null model should assume that all \*individuals\* are equally likely to arrive, not all species; in other words, the relative abundances of species (zOTUs) in the regional species pool should significantly

impact colonization patterns, and thus should affect null expectations about arrival times. Indeed, in some cases the composition of the community itself will partly arise from the relative abundances of taxa in the regional species pool, e.g., if some populations persist or are common simply due to propagule pressure.

**\*\*\* As stated above, we now include an “individual null” model that samples individual observations instead of sampling species. This is very similar to what is used in Raup-Crick or Beta-MNTD analyses, although subsequent analysis of simulated communities is obviously different in our implementation. Per the reviewer’s comment, this “individual” null model uses relative abundance information from a rarefied OTU table. For the “Moving Pictures” data set, the results of the “individual” null model strongly support the results of our original analysis, in that palm and fecal communities exhibited significant underdispersion, but tongue communities did not. For the other two data sets, all communities exhibited significant underdispersion under the “individual” null model, which mostly aligns with what we found earlier. We thank the reviewer for suggesting this alternative approach, because as the reviewer noted, readers may worry that underdispersion may be a conclusion determined by the relative frequencies of species within a data set, and these results address that concern.**

A second assumption of the null/neutral model is that the regional species pool is static over time, even though it likely is not; this assumption isn’t as problematic to me, but it should probably be mentioned.

**\*\*\* We agree with the reviewer that this is worth mentioning. We have included this in our discussion section, next to the aforementioned discussion of regional species pool being pre-underdispersed.**

Overall, I feel the authors need to explore these assumptions, and how they might impact their results. Are we sure that these assumptions do not fundamentally undermine the authors’ main conclusion of underdispersion?

**\*\*\* We found that the “individual-null” corroborates our previous findings. We believe this addition addresses the reviewer’s concerns and improves the robustness of our paper. Discussion of assumptions/limitations has been added as described above.**

A few other minor things:

- I like the added section about detection and recruitment in the Discussion (L330-L343). Indeed, these details seem critical to understand the study. Because of their importance for comprehension, I think this clear description should be present in the Introduction, not the Discussion.

**\*\*\* We agree with the reviewer that this is an important point. We moved our treatment of detection vs. recruitment to our introduction.**

- Related: The shift to using “detection” is an important one, but it has been overdone in my opinion. I think it is important to state early and often that there are methodological limits to our ability to detect

recruitment. However, I think this only needs to be mentioned a few times, and then “recruitment” should be used thereafter. If the paper is ultimately trying to understand recruitment patterns, then it is confusing to not use the word recruitment when discussing hypotheses and results. As it stands, the reader has to mentally translate “detection” to “recruitment” over and over. It leads the reader to repeatedly ask the question: If detection ability is such a problematic experimental limitation to inferring recruitment, then how can we justify using it to understand recruitment patterns?

**\*\*\*We agree with the reviewer that the use of “detection” everywhere is perhaps overdone. We have made the distinction between the two terms clear in our Introduction as advised by the reviewer. Because of this distinction, we still use “detection” throughout much of our Methods and Results sections. But per the reviewer’s advice, we have replaced many instances of “detection” with “recruitment” throughout our Introduction and Discussion sections, where appropriate.**

- It should be made clear in the abstract that overdispersion/underdispersion results in this study refer to the \*order\* of arrival.

**\*\*\*We agree with the reviewer and have made this change.**

Referee #2 (Comments to the Author):

I thank the authors for carefully and thoughtfully addressing all of my major concerns. The authors very clearly modified the manuscript to identify the major limitations and assumptions of the study and to avoid unsupported generalisations. I believe the revised manuscript presents a very strong study and will be of broad interest to those in microbial ecology.

**\*\*\*We hope the current revisions to our manuscript also meet with the reviewer’s approval.**