Dear Editors,

Thanks so much for giving us the opportunity to revise our manuscript. We also would like to thank these reviewers for reading our manuscript so closely and for their insightful and constructive comments. We believe that it has greatly improved the manuscript. A detailed point-by-point response to these comments is included below:

Best Regards,

John L. Darcy

**Responses to reviewers**

**Author responses are in bold and begin with \*\*\*.**  
  
Comments from reviewer(s):   
  
Referee #1 (Comments to the Author):   
  
In this manuscript, Darcy et al. examine the changes in community structure following the arrival of new taxa in a variety of host-associated microbial communities. They ask whether new arrivers are more or less phylogenetically related to their nearest local relative than expected by chance. Overall, I found the study very compelling. I particularly liked the focus on changes in community structure through time with a strong focus on immigration, rather than relying on single snapshots -- a limitation of much empirical community structure work in the past. I was also very impressed with the transparency of the analysis (all code is well organized on github and even comes with a tutorial to understand the analytical approach!), which sets a standard that new submissions should strive to match. However, I have a few major concerns that I would like to see addressed (I think they can be), and a few minor points.

**\*\*\* We are very glad the reviewer found our study compelling, and that the reviewer noticed our code on GitHub. We are especially proud of our tutorial vignette.**   
  
MAJOR CONCERNS   
There has been a lot published about using patterns of community membership (e.g., phylogenetic over- and under-dispersion) to infer community assembly processes in the last 20 years, and the lessons learned over this period are not sufficiently incorporated into the manuscript. For example, while classical literature predicts that competition will lead to trait-based overdispersion, more recent literature suggests that competition can lead to groups of similar species (i.e., closely related species, assuming trait conservatism) (e.g., Mayfield and Levine 2010). Indeed, the same models that predict overdispersion in stable equilibrium predict clusters of similar species leading up to that equilibrium (Scheffer and van Nes 2006, D’Andrea et al. 2019), and ecological forces such as immigration or periodic environments can maintain these clusters indefinitely as a stationary state (Sakavara et al. 2018). In other words, 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. Please re-frame your discussion with a more thorough treatment of current theory.

**\*\*\* Thanks so much for bringing this highly relevant literature to our attention. We have revised both our Introduction and Discussion sections using the citations the reviewer provided (lines 97-101, 345-351, 400-401). We now introduce more contemporary theory that supports our underdispersion hypothesis, and discuss our results in that context as well.**

A second but closely related concern is how “overdispersion” and “underdispersion” are used. Generally, a community with phylogenetic overdispersion is one that is overdispersed at a given point in time, relative to (e.g.) a regional species pool. However, in this manuscript, phylogenetic overdispersion refers (I think) to how phylogenetic structure shifts following the addition of a new set of recruits between two time points. As such, a community can be overdispersed relative to a null expectation based on a hypothetical regional community, but may recruit species in a way that makes it less overdispersed, and therefore it would be tagged as a community that is more overdispersed than expected by chance. (Correct me if I'm wrong here.) I’m not saying “overdispersion” and “underdispersion” can’t be used to frame results results in this study, but I think the authors should be more careful with their usage, and avoid potential confusion with prior usage of the terms by clearly articulating their intended interpretation in the introduction.

**\*\*\* We agree with the reviewer that our use of the words “overdispersed” and “underdispersed” may be different than what some readers may expect. In order to clarify our use of these terms, we have refined our hypotheses in our Introduction to refer to over- or underdispersion specifically as these terms relate to the new species that are detected. We now also include a paragraph in our Introduction detailing how our use of the terms differ from some previous work (Lines 111-120).**

I would have liked a more thorough discussion of the assumptions that go into using \*cumulative\* phylogenetic diversity, rather than simply the phylogenetic diversity of the prior sample. To illustrate my concern, imagine a community in which one particular species is extirpated by competitive exclusion. Then, several time points later, a close relative recruits into the community. With the authors' current method of using cumulative phylogenetic diversity, we would conclude that the species is recruiting into a community with a close relative, and therefore suggests underdispersion. However, in reality, at that time, the species is recruiting into a community without any close relatives, and indicates overdispersion. When examining recruitment success, why pay any attention to the state of the community several time points prior to the current sample, rather than simply to the sample immediately prior? Please justify your current approach and state the assumptions that go into  
interpreting its results.

**\*\*\* We agree with the reviewer that our use of cumulative phylogenetic diversity decreases our ability to link the results to direct interactions between species. Our model is used to investigate the extent to which newly detected species are likely to be similar to previously detected close relatives, but indeed the previously detected species may no longer be present in the community. However, since time-series data are often conducted with very variable spans between sample points, our model is more broadly applicable to sparsely sampled time-series data and has relevance to human microbiome studies, such as by producing an understanding of whether a pathogen’s probability of detection may be higher if a conspecific strain was previously observed, even across a longer time-span. However, our work does also address the question of whether more recent interactions are specifically important by using a sliding-window analysis of 5-day intervals in a dataset with intensively-sampling, thereby making this approach feasible. This sliding window analysis still showed significant underdispersion in a majority of windows analyzed. We now include a paragraph in our Discussion that addresses the cumulative aspect of our model, what our results really mean, and how our model should be interpreted (Lines 356-368). We have edited Figure 1’s caption for clarity as well.**  
  
The first paragraph of the discussion describes the questions well, I think some of the same phrasing could even be repeated in the Introduction. It clearly states that the focus of the study is on the order of arrival, not overall community structure, and this is key.

**\*\*\* We have incorporated that language into our Introduction (Lines 115-120).**

MINOR CONCERNS   
- Sometimes the writing is too informal. For example: don’t refer to intuition (L441), but rather explain initial logic. Don’t use strained analogies (e.g., Burger King, L485). Don’t use terms like “nice-looking fits” (L356)

**\*\*\* We have made these changes.**  
  
- Define the assumptions underlying the term ‘arrival’ in the main text, not in a footnote.   
Don’t conflate the terms “species pool” and “metacommunity” pool. Metacommunity theory comes with specific assumptions that may not apply to your system.

**\*\*\* In response to both reviewers’ comments, we have changed the way we talk about “arrival” in the paper. We now use the word “recruitment” instead for the natural phenomenon under question, since we already used that term in the paper and now feel that “arrival” had too much of a biogeographic connotation. Per the advice of both reviewers, we now are much more cautious when discussing empirical data, which consist of species detections, not actual recruitment/arrival events. This difference is introduced in our Introduction (Lines 71-74) and discussed in our Discussion (Lines 331-344). We have removed references to the species pool as a proxy for the metacommunity.**   
  
- I don’t think it is necessary to mention Darwin’s pre-adaptation theory more than once in the introduction. Stick with more recent theoretical predictions.

**\*\*\* We actually only mentioned Darwin’s pre-adaptation theory** **once in the Introduction but contrasted it with Darwin’s naturalization hypothesis. We do still think that this is appropriate to include but following this reviewer’s suggestions, we have revised both our Introduction and Discussion sections to put our work into the context of more recent theoretical predictions (lines lines 97-101, 345-351, 400-401).**   
  
- I think the study is OK without considering species abundances, but I think you should at least mention how species abundances could affect results, and how you might incorporate abundances into future work. The word “abundance” is not mentioned in the manuscript, which strikes me as an oversight.

**\*\*\* We agree with the reviewer. We now discuss how abundance data could potentially be incorporated into our model in future studies (Lines 339-342).**  
  
- Just because underdispersion is commonly observed in hosts it doesn’t mean it is optimal for probiotics or promoting host health (L488)

**\*\*\* We fully agree with the reviewer, and we didn’t mean to imply that it was optimal. We have revised this section accordingly.**

Referee #2 (Comments to the Author):   
  
This is a fantastic manuscript that attempts to measure priority effects and community assembly processes from time-series data. The authors develop a creative and original statistical model to test whether new immigrants into a community are more or less phylogenetically related to prior immigrants than would be expected by chance. This allows one to assess key processes involved in community assembly, such as the relative importance of competition/exclusion and pre-adaptation. The model can therefore potentially be used to address longstanding and timely question in microbial ecology, and I commend the authors for their efforts. The manuscript is also very well written, clear, and concise. While I am generally positive about the manuscript, I nevertheless have several serious concerns.

**\*\*\* We thank the reviewer for their positivity and for their thoughtful review.**

1) My main concern is one that the authors have certainly heard before. The authors use the model to predict probabilities of immigration events based on the contemporary composition of the community. The main problem is that the empirical datasets are time-series datasets that do not provide evidence of immigration events. They simply show when a particular species exceeds the analytical detection limit of the applied methodology. There is no way of knowing whether the detection of a species represents a true immigration event or not. To acknowledge this limitation, the authors included a footnote below line 148 stating that this limitation has nothing to do with the model but instead is due to analytical and experimental design issues. This footnote is convincing, but in my opinion misses the point. I have three major issues with this:

**\*\*\* Both reviewers asked for this issue to be more thoroughly discussed, but Reviewer 2’s points below were especially helpful in addressing this. We respond to each sub-point below.**   
  
+ The text continuously talks about immigration events with regards to these datasets, even though these may not be immigration events as acknowledged in the footnote. For readers that may not be interested in the methodological aspect, this could cause confusion in the field. I would encourage the authors to remove the terminology “joining the community” and instead use more cautious terminology such as “analytically detected”.

**\*\*\* We have followed the reviewer’s advice. This comment caused us to evaluate our word choice in the paper, and we now realize that our language (“arrival”, “recruited”, “joining the community”) was variable and unclear. We now use the word “recruitment” to describe the natural phenomenon under question. The reviewer is right to be cautious about terminology, especially in a paper as technical as ours. We agree that readers who skim or skip our complex methods section may have been confused by our previous draft. We now discuss “species detections” as the empirical consequence of recruitment, and we make it clear that we are actually modeling these detections. This concept is now introduced in our Introduction (Lines 71-74) and made even more clear in our Discussion (Lines 331-344). We have edited our whole manuscript so that we use “detected” whenever we discuss our model and empirical data, and only discuss “recruitment” as the natural phenomenon that results in detection. We use “recruitment” in the title, because it is the goal of our model to understand that process, and because “detection” makes the paper sound like we are discovering new species. But we now clarify that the model is used on empirical detections in our abstract as well (Lines 38-43).**  
  
+ While this concern is rooted in analytical and experimental design aspects rather than the model, a major question remains unresolved. What type of datasets and experimental design would be sufficient to draw the conclusion that the authors want to draw? My gut tells me that such datasets don’t exist. If this is the case, then what use is the model? This needs discussed.

**\*\*\* We have revised our manuscript to address this issue in two ways. First, as the reviewer suggested, we now discuss empirical species detections throughout our manuscript. In this way, it is more clear that our model is used to describe data. Second, we now discuss the extent to which microbiome data reflect actual species presence/absence (lines 332-339). We also now discuss how future experimental designs could help to alleviate this issue (lines 339-344).**  
  
+ This entire topic is missing from the discussion, and is instead confined to a single footnote. This seems inadequate for such an important point (at least it seems important to the naive reader). I would encourage the authors to expand on this in the discussion and (perhaps) even in the introduction.

**\*\*\* We now introduce this topic in our Introduction (lines 71-74) and discuss it in our Discussion (Lines 331-344). We also allude to it in our abstract (lines 38-43), just in case that is all that a reader sees.**   
  
2) The model was applied to a very limited number of independent experiments, which is evident by the incredibly short results section (merely 3 paragraphs). This precludes the possibility for the authors to draw any general conclusions. Nevertheless, the authors try to make general conclusions, such as underdispersion being a general trend for the human gut and potentially skin microbiomes (e.g., lines 483-485), I find this reckless given the small number of datasets. If the authors would like to draw general conclusions (and I certainly think that’s possible), I would suggest they apply their model to a far larger number of datasets.

**\*\*\* We have removed those generalizations from our manuscript, and we thank the reviewer for encouraging our future work.**