Responses to comments on "Community size affects the signals of selection and ecological drift on biodiversity".

Tadeu Siqueira, on behalf of all authors

April, 2019 – Rio Claro, Brazil

The following commentaries were made by recommender Dr. Eric Harvey. We provided responses to these commentaries (in bold) where we felt it was appropriate.

Dear Dr. Siqueira,

Thank you very much for submitting your manuscript "Community size affects the signals of selection and ecological drift on biodiversity for recommendation" at PCI Ecology. The reviewers and I appreciate the work you have accomplished. Based on the reviews, we will not be able to recommend this manuscript at this point, but will be happy to consider a revised version.

Both reviewers and I agree that this manuscript tackles a topic relevant to re-searchers interested in stream ecology as well as, more broadly, to community ecologists. In particular, I think that the contrast between boreal and tropical systems is very interesting as it provides potential explanations for the idiosyncrasies observed in the many studies addressing this issue at different specific locations.

However, both reviewers raise concerns, mainly about some of the assumptions behind the null model approach, and the interpretation of the results. Those issues should be addressed clearly in the manuscript.

I also agree with one reviewer that the fact that the data is only described in Heino et al., 2018 leads to some unclarity in the Methods section that should be addressed. Both reviewers and myself found that the different approaches used to define community size need to be clarified. Another important issue is with expectations and interpretations of the slope of the relationship between beta-deviation (and beta-diversity) and community size. The authors state clearly predictions for a beta-deviation of 0 versus 1 or -1 but they do not provide clear interpretations for the slope itself (positive versus negative). One reviewers made suggestions to help with that. This would greatly improve the clarity since they found that with one metric the slope is positive and with the other one the slope is positive. Finally both reviewers suggest complementary analyses that should be considered to clarify those issues.

Should you decide to revise the manuscript for further consideration here, your revisions should address the specific points made by myself (this email and minor comments attached with this email) and each reviewer. Please include a cover letter indicating your responses to the review comments and the changes you have made in the manuscript. If you disagree with a reviewer's point, explain why. Also, please add line numbering to the manuscript so that it is easier to refer to specific lines.

Sincerely, Dr. Eric Harvey Recommender, PCI Ecology

Dear Dr. Eric Harvey,

Thank you very much for handling our manuscript and for the opportunity to submit a revised version. We are glad you think our manuscript addresses a relevant topic that goes beyond the particularities of stream systems. You and the reviewers provided constructive commentaries and well-argued points that we much appreciated. We responded to all questions raised by you and the reviewers and addressed most issues. When the issues were not addressed in the revised manuscript, a full explanation was provided. We hope that this revised manuscript could be considered for recommendation in PCI Ecology. Of course, we are completely open to new suggestions and comments, and we will consider any new amendments if they are necessary.

Minor comments

Introduction
Recommender's annotations on the introduction "Demographic stochasticity can override ecological selection in small populations leading to the co- existence of strong and weak competitors.

Importance of deterministic vs. stochastic processes on beta-diversity -> Environmentally similar local communities may differ in species composition due to distinct legacies of demographic stochasticities —> human perturbations re- duce size of communities and thus might increase the effect of demographic stochasticity (and influence beta-diversity in a predictive way)."

Comments. I really appreciate the introduction and the general context of the study. I think that it is very clearly laid out with appropriate literature cited.

R. Thank you for these positive comments.

MAIN OBJECTIVES:
"In this study, we tested the hypothesis that ecological drift is a major process causing variation among small communities. We expected that ecological drift would play a smaller role in large communities where deterministic niche selection should drive spatial variation in community structure."

Predictions
"we expected that beta diversity would be high and beta deviations would be close to zero in watersheds with the smallest communities (some watersheds in Brazil only)."

"Second, we expected that watersheds with larger communities in Brazil would have lower values of beta diversity compared to smaller communities, but high positive values of beta deviation."

"Together, these two predictions would lead to a negative relationship between beta diversity (before controlling for sampling effects) and community size, but a positive relationship between beta deviation and community size."

"Finally, because the smallest boreal stream communities are as large as the largest tropical communities (Heino et al. 2018), we expected that boreal communities would show a weak or lack of relationship between (positive) values of beta deviation and community size."

Comments. The tropical vs. boreal contrast is very interesting. Suggest that there are large-scale latitudinal gradient in the relative importance of stochastic and demographic processes.

R. Thanks. Because our data does not include a true latitudinal gradient (but two different regions in contrasting latitudes), we decided not to suggest there could be a latitudinal gradient in the relative importance of stochastic and demographic processes. But this is certainly an interesting point that deserves further investigation.

Methods
"...and estimated local community size as the mean number of individuals sampled in a watershed. Because streams within and among regions differ in width and this could be viewed as measure of habitat size, we multiplied local community size by stream width, averaged it within watersheds, and defined it as an alternative measure of community size."

Comments. I am confused: if community size is a property of the watershed how can it be multiplied by each within watershed locality stream width?

R. Sorry, the text was not precise there. Actually, following Orrock and Watling (2010), community size is a property of each community (i.e., a stream site in our study). However, our response variable was beta diversity, i.e., variation in species composition among communities, which was measured as the mean of a compositional dissimilarity matrix between all pair of streams within watersheds. Specifically, for each watershed, we have 5 streams. Thus, we needed our explanatory variable (community size) to correspond to the spatial scale at which we measured our response variable. That's the reason we averaged the number of individuals (collected at each stream) across five streams within each watershed. In short, we need a measure of abundance for each watershed. Following this strategy, both beta diversity and community size had 20 values (one per watershed; i.e., one per metacommunity), estimated at the same spatial scale. We modified the text to accommodate this (L. 189). Now it reads: "*We adopted the definition of community size provided by Orrock and Watling (2010) and estimated local community size as the number of individuals sampled in a stream site. However, as β-diversity was estimated at the watershed scale (i.e., considering 5 stream sites; see below), we averaged the number of individuals across five streams within each watershed. This resulted in 20 values of community size, one per watershed (Fig. S1)*". We also created a new Figure (see below) to explain these procedures.

Basic model tested: β = *f* (CS,SE,EH)
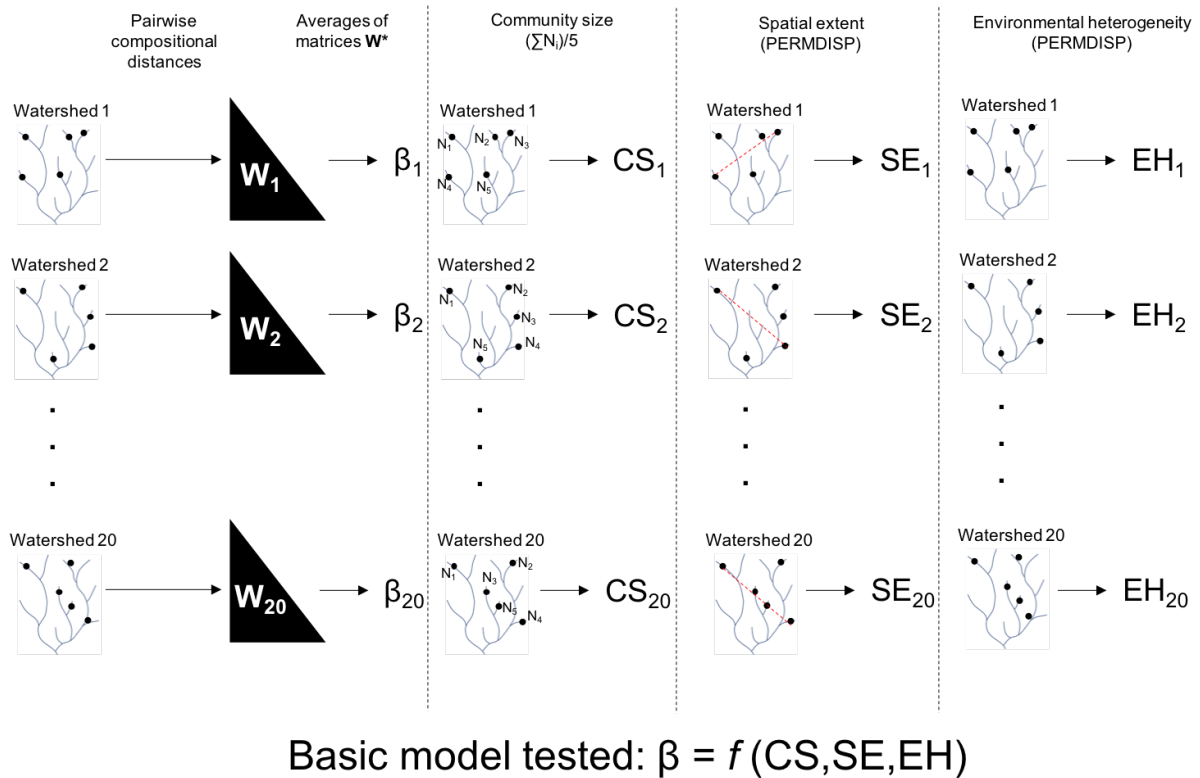
* This was repeated for different dissimilaity coefficients, and for null model analyses

Figure S1. Graphical representation describing the general procedures for estimating beta diversity (β), community size (CS), spatial extend (SE) and environmental heterogeneity (EH) within watersheds (n = 20) in Brazil and Finland.

Comments. Also, each watershed is considered a metacommunity, right? Here my understanding is that local community size is averaged across the whole meta community?

R. Yes. We did this because we wanted one estimate of beta diversity and community size per metacommunity. The abundance in each stream (N_i) was given by the number of individuals in a stream. After, the abundance in each watershed (metacommunity) was given by the $\sum N_i$/number of streams. The text was modified to make this clear (L. 190).

Comments. Are different stream orders equally represented in each watershed, or did you only sample a certain range of steam orders, or was it at random?

R. They were generally of the same order within each watershed, but varied a bit among watersheds, including 2nd and 3rd order streams in Brazil, and a few 4th order streams in Finland. We included this information in the manuscript (L. 171).

Comments. for the median size approach; do you mean median community size rather than median population size?

R. We mean median number of individuals across different genera. For example, in a given stream gen.1 = 3 individuals; gen.2 = 5 individuals; and gen.3 = 7 individuals. The

median = 5. We did this for the five streams within a watershed and averaged the medians. We noted that both reviewers also thought this and other definitions of community size were confusing. Because results were very similar no matter the definition we adopted, we decided to remove theses sentences from the main text. Thus, community size is simply the number of individuals per stream, averaged across 5 streams within a watershed.

"Fitted models provided similar results with all measures of community size and, thus, we show here results based on the former measure. We repeated the procedures described above but changing the definition of species pool to the watershed scale (not the entire region; step (ii)). "

Comments. IF all values are averaged at the watershed/metacommunity scale it's unclear how beta-deviations can be measured at the within watershed scale?

R. First, let's consider our incidence-based measure of beta diversity. Within a given watershed, we estimated the Sørensen dissimilarity coefficient between all pairs of streams and averaged them so that we had one unique value of beta diversity for that watershed. Please, keep in mind that there were 5 streams per watershed (see figure below). To estimate beta-deviation for that same watershed, the algorithm creates new communities for each stream in that watershed by sampling species (or individuals; it depends on the dissimilarity coefficient) from the regional pool (all 100 streams in a region). The algorithm stops when each stream reaches the same richness (or abundance) as the observed stream. Once all five communities were randomly assembled, the algorithm estimates the Sørensen dissimilarity coefficient between all pairs of streams. These last two steps are repeated 10000 times. Then, the algorithm averages the 1000 values of each dissimilarity pair so that we have one mean dissimilarity value, under random assembly, for each pair of streams, and finally averages these pairwise values within each watershed. The final step consists of subtracting this 'within watershed mean random value' from the 'observed value'. This results in one unique value of beta deviation per watershed, which represents how the observed beta diversity differs from a random beta diversity. We tried to clarify this issue in the main text (L. 197).

Comments. The description of the PERMDISP suggests that the analysis was performed within watersheds. This is confusing with the information provided before stating that beta-diversity metrics were averaged at the watershed scale.

R. We modified the text to clarify that all analyses were done in order to obtain estimates of watershed level variables. We had 20 watersheds, so we need 20 values of beta diversity, beta deviation, community size, environmental heterogeneity, and spatial extent. These 20 values were used in our linear models:

beta deviation ~ community size + environmental heterogeneity + spatial extent.

Hope figure S1 clarifies this issue.

Figures and information in the results section suggest that the PERMDISP was not performed at the same scale as the information in the figures, is that correct?

R. No, it is not correct. They were performed at the same scale as the information in the figures. PERMDISP is a method that can be used to estimate the mean distance of a number of objects projected into an ordination multivariate space to their centroid. Here, we used PERMDISP to estimate the mean distance of five streams to the watershed centroid. We did this to all watersheds so that we ended up with 20 values of environmental heterogeneity (one per watershed), and spatial extent (one per watershed).

RESULTS
Comments. Figure 1. - very interesting! So analysis are done among watersheds and not within watersheds?

R. Yes. We were interested in variation among watersheds, i.e., among metacommunities. But estimates for each watershed were done using averaged within-watershed information. To make our analytical pipeline clear, we decided to produce the Figure shown above.

DISCUSSION
"Mechanistic explanations for the major role of ecological drift in small communities involve the alteration of competitive outcomes of species with different fitness "

Comments. This is a consequence but not a cause of the importance of ecological drift, right? The sentence is unclear.

R. Based on previous findings by Orrock and colleagues, the rationale here is this: In small communities, demographic stochasticity can change the expected outcome of the interaction between a stronger and a weaker competitor. This would cause communities to drift – to fluctuate randomly. In that sense, demographic stochasticity is the cause (it changes competitive outcomes) and drift is the consequence. In our study, however, we cannot assume that competition is the only mechanism involved. So, we decided to be more general and changed the sentence to (L. 330): "*Explanations for the major role of demographic stochasticity in small communities involve the alteration of the occupancy frequency and relative abundance of species with different fitness (Orrock and Watling 2010). When local communities are small, even species with high fitness are at a high risk of extinction due to demographic stochasticity in comparison to a situation when communities harbor large populations. Consequently, species with low relative abundances have a chance to increase their populations in small communities (Orrock and Watling 2010, Gilbert and Levine 2017)*".

Comments. The results suggest that smaller communities should have higher beta-diversity but also higher local richness compared to larger communities dominated by a few species. I apologize for the self-advertisement (I generally avoid to do this), but in that case I feel like this recent study would be very relevant to cite: "Harvey Eric, Gounand Isabelle, Fronhofer Emanuel A., and Altermatt Florian. 2018. Disturbance reverses classic biodiversity

predictions in river-like landscapes. Proceedings of the Royal Society B: Biological Sciences 285:20182441." - but I will leave this at the authors discretion (this is only a suggestion)

R. That was not the case. Smaller communities should have lower species richness and higher beta diversity if there are mainly driven by drift. The point here is that each locality has a different set of (reduced) species composition, making beta diversity higher. Also, with low richness, even one local extinction of species affects pairwise dissimilarities strongly while does not so in high richness. In larger communities, local species richness is higher, but a set of species with high fitness occur in most of the localities, making beta diversity lower.

Comments. So if I understand well, you observed a saturation along the latitudinal gradient for species occurence but not for species relative abundance, correct? This might indicate that variations in relative abundances are intrinsically more stochastic?

R. We are not sure if we understand the comment about saturation. But regarding variation in relative abundances, our results suggest the opposite – when we included variations in species relative abundances, we found beta differed from what would be expected under stochastic assembling; i.e. beta deviation values were always positive and far from zero. This means that including species abundances modifies our perception of the importance of niche selection.

Thank you for your comments.

The following notes were made by reviewer Dr. Romain Bertrand.

In this manuscript Tadeu Siqueira and colleagues test the hypothesis that small communities are more dissimilar among each other because of ecological drift than large communities, which are mainly structured by niche selection. They investigate this issue from in situ ecological observations (combining biodiversity and environmental data from boreal and tropical streams). They used linear models to test the effect of community size on beta diversity, incidence- and abundance-based beta deviations (that are metrics comparing observed beta diversity to null expectations). As expected from theory and recent experimental evidence, they demonstrate that small communities are more driven by random processes than large communities. As a consequence, the authors suggest that ecological drift plays an important role in small communities by increasing the chances of species with low competitive ability to occur within the metacommunity. They conclude that environmental pressures will make smaller communities more vulnerable to novel conditions and community dynamics more unpredictable, as random demographic processes should prevail under these conditions.

The article is well written and results are original and of interest for community ecology. However I have two main concerns which can lead to conduct more analysis, clarify or revise some assumptions, interpretations and conclusions. The most important concern is about expectation and interpretations of deterministic drivers underlying beta diversity, but

it does not challenge the main result of the study, that is beta diversity in small communities is more driven by random processes than in large communities.
All the best, Romain Bertrand

R. Thank you very much for your detailed comments. They were all useful to improve the quality of our manuscript. Below, we provide detailed responses to each of the comments.

Main concerns:
1) Assumptions and interpretation about deterministic drivers underlying beta diversity:
- Tested expectation/assumption number 2: "Second, we expected that watersheds with larger communities in Brazil would have lower values of beta diversity compared to smaller communities, but high positive values of beta deviation. This would indicate that niche selection and sufficient dispersal rates are the main processes resulting in large communities to be more dissimilar than expected by chance."
I don't have the same interpretation that the authors. When beta deviation is greater than 0, ecological or biological processes lead to higher community dissimilarity than expected by chance only (Chase et al. 2011). The main deterministic drivers that can explain such a pattern are niche selection inside the focal area (the watershed in the present study) and when dispersal among sites is low (leading to dispersal limitation). So why the authors consider that sufficient dispersal rate lead to dissimilar communities? May be it's just an error and authors wanted to say "insufficient"?

R. We partially agree. Positive values of beta deviation indicate that community dissimilarity is higher than expected by chance as you said, which further suggests a major role of deterministic niche selection. However, theory and evidence suggest that strong species sorting through niche selection occurs when dispersal is sufficient to allow individuals reach sites that match their ecological requirements (Leibold and Chase 2018). If dispersal is insufficient (i.e., limited), individuals might not reach those sites. This combination of limited dispersal and niche selection results in spatial patterns predicted by the "patch dynamics archetype" (Winegardner et al. 2012; Leibold and Chase 2018). We recognize that limited dispersal could have also generated the observed patterns (positive beta deviations). However, the watersheds (metacommunities) we studied are not large enough to lead to strong dispersal limitation. In Brazil, maximum distances between pairs of streams within watersheds varied from 2.48 to 8.86 Km, whereas in Finland it varied from 12.77 to 109.5 Km.

We modified the sentence to clarify this issue (L. 122) and added a figure to represent our expectations: "*Second, all else being equal, we expected that (E2) β-diversity in watersheds with the largest communities in Brazil and Finland would be far from null expectations, but lower than in the smallest watersheds (Fig. 1). This would indicate that niche selection and sufficient dispersal rates are the main processes resulting in large communities to be more dissimilar than expected by chance, as species sorting occurs when dispersal is sufficient to allow individuals to reach sites that match their ecological requirements (Winegardner et al. 2012, Leibold and Chase 2018). Although dispersal limitation can also cause positive values of β-deviation, this was unlikely in our study system as the watersheds we studied are not large enough to lead to strong*

dispersal limitation (see details below). Taken together, these two expectations would lead to a negative relationship between β-diversity (before controlling for sampling effects) and community size, but to a positive relationship between β-deviation and community size (Fig. 1)."
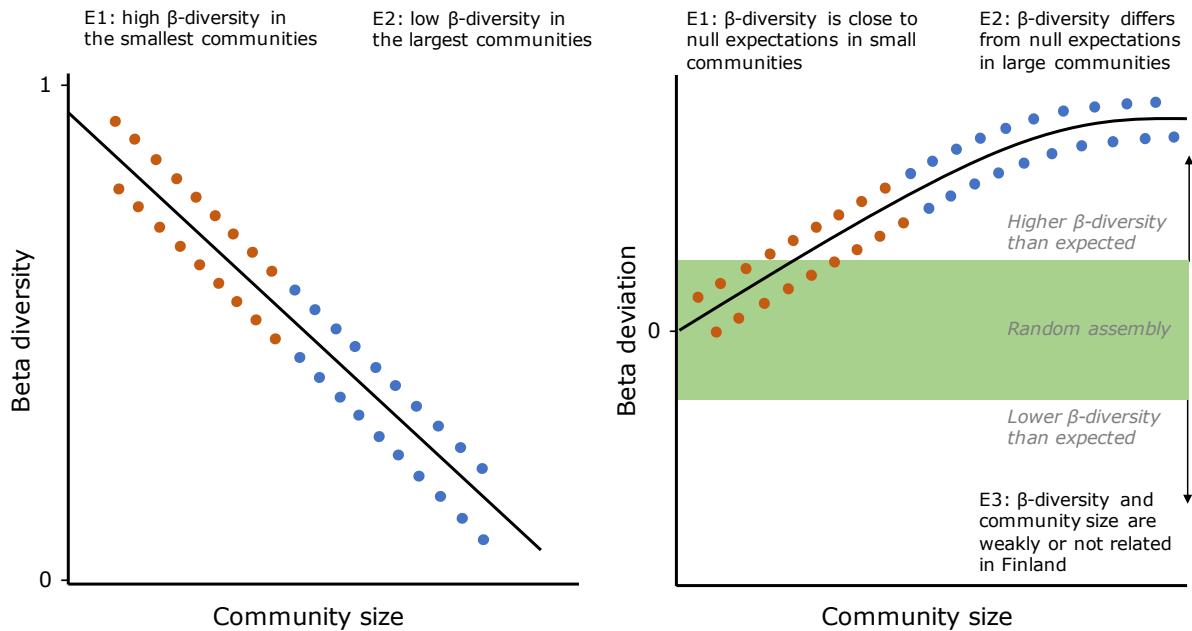

Figure 1. Graphical representation describing a priori expectations (E) about the relationship between beta diversity and beta deviation with community size in Brazil and Finland.

- Tested expectation/assumption number 3: "Finally, because the smallest boreal stream communities are as large as the largest tropical communities (Heino et al. 2018), we expected that boreal communities would show a weak or lack of relationship between (positive) values of beta deviation and community size. This would indicate that deterministic niche selection is the main processes determining the structure of boreal communities because boreal stream insect communities are usually large enough and not subject to strong demographic stochastic effects."

I don't really get the justification of this expectation. It's like the authors expect a non linear relationship between beta deviation and community size, with a community size effect decreasing as increase the community size. Authors have to explain a bit more their expectation and used previous works to justify their choice.

R. Indeed we expected a non-linear relationship (which was not confirmed by our results). Using mechanistic models, Orrock and Watling (2010) showed that the proportion of patches occupied by strong vs weak competitors varies non-linearly as community size changes (see Fig. 2 in Orrock and Watling 2010). Superior competitors occupy a constant low proportion of patches (close to zero or zero) along a gradient of community size – from very small communities to mid-sized communities. From that point, there is a linear increase in patch occupancy by strong competitors (and a linear decrease in occupancy by weak competitors) until another point where it stabilizes again (i.e., where communities are large). From that point on, further increases in community sizes do not alter

competitive outcomes. In our study, we expected that an interaction between ecological drift and niche selection would occur until a point where community size was large enough to deterministic niche selection reach its maximum strength. That would occur in large-sized Finnish communities. We changed the text to make the text clearer in this respect (L. 133): *"Finally, the mechanistic model by Orrock and Watling (2010) suggests a non-linear relationship between metacommunity dynamics and community size. Superior competitors should occupy a constant low proportion of patches (close to zero or zero) along a gradient of community size – from very small communities to mid-sized communities. From that point, there should be a linear increase in patch occupancy by strong competitors (and a linear decrease in occupancy by weak competitors) until another point where it stabilizes again (i.e., where communities are large). From that point, further increases in community sizes should not alter occupancy frequencies as community size would be large enough to allow deterministic niche selection to reach its maximum strength. Thus, because the smallest boreal stream communities are as large as the largest tropical communities (Heino et al. 2018), we expected that (E3) β-deviation and community size would be weakly or not related in Finland (Fig. 1)"*.

Furthermore, is species diversity the only factors which could explain such a pattern in the present study? I mean authors compared two different regions with contrasted community size that do not overlap at all. So how to be sure that increase in community size is the only factor which can explain a weak or lack of relationship between beta deviation and community size? It could the difference in environmental condition between region. To properly tested for that authors should investigate the relationship between beta deviation and community size within a region with large range of community size (from low to very large community).

R. We are not sure if we understand this suggestion properly because, as suggested by you, we indeed tested the relationship between beta deviation and community size within each region (please, see Fig. S1, Table 1 and Figure 3). Also, in both Brazil and Finland, our sampling scheme was purposely designed to comprise a large environmental gradient both in terms of watershed land use and in-stream characteristics. Thus, we are not sure we would be able to find even larger communities in Brazil or smaller communities in Finland. But we are sure about one thing: community size varies a lot in both regions: Brazil (range: 12 to 488 individuals per stream); Finland (range: 20 to 3630 individuals per stream).
Regarding confounding variables, we did include environmental heterogeneity and spatial extent in our models, as this is what theory suggests should matter for beta diversity.

-In the introduction, authors explained that "we need estimates of beta diversity that account for differences both in species richness and species relative abundance". So authors presented results for incidence- and abundance-based beta deviation, and demonstrated opposite relationships of these metrics with community size (and contrasting effects of environmental heterogeneity). But they did not discuss these differences.

R. The second, third and fourth paragraphs of the discussion section were entirely dedicated to explaining why we found that (1) beta deviation was close to zero no matter

the dissimilarity coefficient; (2) incidence-based beta deviation was negatively related to community size; and (3) abundance-based beta deviation was positively related to community size. We have included some sentences in the last paragraph of the discussion section to summarize the main differences (L. 443). We also provide more details about this issue below in response to another comment.

"The magnitude of β-diversity deviation from the null models (negative and positive values) and how it related with community size (negative and positive slopes) indicate that stochastic and deterministic processes affect species occurrences and their abundances differently. While incidence-based β-deviation was negative and decreased with community size in Brazil, abundance-based β-deviation was positive and increased with community size in both regions. These results indicate that: (1) as communities get larger, demographic stochasticity plays a less important role and excessive dispersal combined with niche selection tend to homogenize the genus composition of larger communities; and (2) variations in genera relative abundances are the result of local environmental filtering. However, we acknowledge that these are, at least partially, a posteriori explanations as we expected that the results would be similar (i.e., a positive relationship between beta deviation and community size) independently of the type of coefficient. On the other hand, these contrasting results appear interesting avenues for future research."

Moreover, the authors described general expectations in the introduction which did not depend to the kind of beta deviation investigated. So it strengthens the intriguing characteristic of the results.

R. We did that because we had no a priori expectations (based on theory) regarding possible differences between incidence- vs. abundance-based beta deviation before we ran the analyses. We thought results would be similar – a positive relationship between beta deviation and community size (please, see our reply above).

Why incidence-based beta deviation is negative and increase (at least in tropical regions) with community size while abundance-based beta deviation shows the opposite?

R. This was the most intriguing result and may be the one that advances our understanding in the field. We think incidence-based beta deviation was negative (communities were less dissimilar than random expectations) because: i) environmental gradients within watersheds were not strong enough to produce major changes in the genus composition among streams at watershed scale; and ii) the "long-term availability for colonization and multivoltinism allowed some genera to reach widespread distribution within tropical watersheds (Saito et al. 2016). Similarly, good dispersal abilities of northern species exceeded the role of within-watershed environmental heterogeneity in the boreal region, making these communities less dissimilar than expected".

We think incidence-based beta deviation decreased with community size in Brazil because smaller communities were more influenced by demographic stochasticity than larger communities. That is, smaller communities had negative values of beta deviation that were close to zero (zero indicates what would be expected under random assembly), while larger communities had more negative values. As communities get larger,

demographic stochasticity plays a less important role and excessive dispersal plays a major role (homogenizing larger communities). We added sentences to make this point clearer (L. 408): "*Thus, as communities become larger, demographic stochasticity plays a less important role, while dispersal surplus plays a major role in homogenizing the genus composition of streams within tropical watersheds. However, this, as almost everything in ecology, should be scale dependent (Chase et al. 2018). If our watersheds were larger (in extent), dispersal limitation would likely have played a role and niche selection could have been the major driver of community structure.*"

We think abundance-based beta deviation was positive because patterns of abundance distribution among genera were more dissimilar than patterns predicted by random assembly. This indicates that niche selection was likely sufficient to cause non-random variations in genera relative abundances among communities – i.e., in terms of which genera were more abundant and less abundant. The relationship between abundance-based beta deviation and community size was positive because as communities get larger, demographic stochasticity plays a less important role. Larger communities within the same watershed differed more strongly from each (than expected by chance) than smaller communities. We acknowledge, however, that these are, at least partially, a posteriori explanations as we expected that the results would be similar (i.e., a positive relationship between beta deviation and community size) independently of the type of coefficient. Please, check the sentences we add to the new version. We hope we have clarified this issue.

We have included a new (third) paragraph (L. 352): "*In general, the positive relationship between abundance-based (Bray-Curtis) β-deviation and community size in both regions is in line with our expectations. The positive slope and values of β-deviation indicate that the variation in community structure among streams increased with community size more than what was expected under random assembly. Thus, in terms of which genera were more abundant or less abundant, and more aggregated or less aggregated, communities within the same watershed differed from each other more than expected by chance, especially in mid to large communities. As dispersal within watersheds was likely not limited, this positive relationship indicates that niche selection was sufficient to cause non-random variations in genera relative abundance and aggregation patterns among large communities. We suggest that as community size increases, demographic stochasticity becomes less important and selection determines which species are more abundant locally and widely distributed within the metacommunity. In this case, small random variations in the number of individuals of relatively abundant genera occurring in larger communities, which can only be detected with abundance-based β-diversity metrics, do not result in major changes in genus occurrence.*"

And some sentences at the concluding paragraph (L. 443): "*The magnitude of β-diversity deviation from the null models (negative and positive values) and how it related with community size (negative and positive slopes) indicate that stochastic and deterministic processes affect species occurrences and their abundances differently. While incidence-based β-deviation was negative and decreased with community size in Brazil, abundance-based β-deviation was positive and increased with community size in both regions. These results indicate that: (1) as communities get larger, demographic stochasticity plays a less important role and excessive dispersal combined with niche selection tend to homogenize*

the genus composition of larger communities; and (2) variations in genera relative abundances are the result of local environmental filtering. However, we acknowledge that these are, at least partially, a posteriori explanations as we expected that the results would be similar (i.e., a positive relationship between beta deviation and community size) independently of the type of coefficient. On the other hand, these contrasting results appear interesting avenues for future research."

Which results we have to consider in priority or are the most robust to explain beta diversity variation? May be both analysis because they provide complementary information...

R. We think both analyses are important as they are complementary. Actually, this was already suggested by Marti Anderson and colleagues (see reference below). They suggested that ecologists should use different dissimilarity coefficients when analyzing beta diversity, as they provide a continuum in beta diversity measures, from a strong emphasis on pure compositional dissimilarity (incidence-based coefficients; e.g., Sørensen and Jaccard) to a strong emphasis on relative abundance dissimilarity (e.g., Bray-Curtis to Manhattan (Anderson et al. 2011; Navigating the multiple meanings of β diversity: A roadmap for the practicing ecologist. Ecol. Lett. 14: 19–28.).

Moreover, sometimes authors mixed results of incidence- and abundance-based deviation to validate their expectations. The authors expected high beta deviation values in large subtropcial communities, and a positive relationship between beta deviation and community size. They verified this expectation for abundance- based beta deviation only. Then, the authors expected a weak or lack of relationship between values of beta deviation and community size in boreal streams. In this case, they verified this expectation for incidence-based beta deviation only (I also noted that in this case, the authors expected such a pattern for positive beta deviation values but here authors reported negative values...). Why considering incidence-based beta deviation to validate some expectation and abundance-based beta deviation in other case?

R. We expected beta diversity would be negatively related to community size. More specifically, we expected that beta deviation would be close to zero in watersheds composed of small communities, whereas beta deviation among large communities would be higher than zero (positive). Note that we did not have a priori expectations for a possible difference between incidence- vs. abundance-based beta deviation. We now include a figure to represent our expectations.
      Please, also note that both incidence- and abundance-based coefficients confirmed our a priori expectations that (1) beta diversity would be higher among small communities and (2) beta deviation among smaller communities would be closer to zero in Brazil. So, in this case, we are actually not "considering incidence-based beta deviation to validate some expectation and abundance-based beta deviation in other case".
      The only result that was a bit surprising was the negative relationship between incidence-based beta deviation and community size. However, this is a valid result that makes sense, as we explain in the third and fourth paragraphs of the discussion section.

- What factors drive the increasing or decreasing relationship between beta deviation and community size? In case of abundance-based beta deviation, authors considered that "niche

selection was likely sufficient to cause non-random variations in genera relative abundances among large communities" because positive beta deviation values increase with community size.

R. Indeed. That is what we think caused beta deviation values to be positive in watersheds composed of large communities. More specifically, we think environmental filtering makes genera relative abundances to vary among mid to large communities more strongly than among small communities. We are not certain about the specific mechanisms behind the positive values of beta deviation (e.g., predation, competition), as our data does not allow such inferences.

In case of incidence-based beta deviation, authors considered that "large size of boreal communities allowed niche selection to be strong enough to drive spatial variation in genus composition among communities" because negative beta deviation is unrelated to community size. Two different results but same interpretation. Why? In the last case, authors found a positive relationship with spatial heterogeneity which suggests niche selection. But how authors can explain that niche selection occurs while beta deviation values are negative ? It would mean that niche selection structured the difference in species composition among communities but that high dispersal rate has a higher effect leading to more similar community than expected by chance. If authors have this interpretation, please write it clearer.

R. This is a very good point. First, you are right, incidence-based beta deviation was mainly negative in both regions. So, the interpretation should be the same – i.e., communities were less dissimilar in genus composition than expected by chance. In other words, they were more similar than expected. The weak positive relationship between (negative) values of beta deviation and environmental heterogeneity is not straightforward to interpret, though. This relationship indicates that higher values of environmental heterogeneity are related to less negative values of beta-deviation. This is tricky as higher environmental heterogeneity would be associated to beta deviation values that are closer to zero or with a tendency to become positive. We rephrased those sentences to make this point less confusing (L. 377):

"*Within watershed environmental heterogeneity was weakly and positively related with β-deviation only in Finland, where community sizes are on average five-fold larger than in tropical streams (Heino et al. 2018). This suggests a tendency towards environmental determinism; the higher the environmental variation within watersheds, the higher their β-deviation. However, our sample size does not allow making good predictions in this regard. Thus, it is likely that even the large size of boreal communities was not sufficient to allow niche selection to be the main driver of spatial variation in genus composition among those communities.*"

In the following paragraph, we discussed the role of high dispersal rates in more detail and, as you suggested, made that point clearer (L. 406):

"*Similarly, good dispersal abilities of northern species may have masked the role of within-watershed environmental heterogeneity in the boreal region, making these communities*

less dissimilar than expected. Thus, as communities become larger, demographic stochasticity plays a less important role, while dispersal surplus plays a major role in homogenizing the genus composition of streams within tropical watersheds. However, this, as almost everything in ecology, should be scale dependent (Chase et al. 2018). If our watersheds were larger (in extent), dispersal limitation would likely have played a role and niche selection could have been the major driver of community structure."

Moreover, "if environmental filtering is strong, sites with similar environmental conditions should be more similar than expected, while sites with dissimilar environmental conditions should be less similar than expected. Likewise, when dispersal limitation is strong, nearby pairs of sites will be more similar than expected, whereas distant pairs of sites will be less similar than expected" as explained by Chase et al (2011). So why environmental heterogeneity has a significant and positive effect on beta deviation in only one case while authors seems to consider that niche selection is a strong driver of community dissimilarity?

R. We are not certain about this. As you pointed, environmental heterogeneity was a significant explanatory variable of beta deviation only for the incidence-based coefficient estimated with the boreal data. Possible explanations for this include: (1) only the beta deviation of the largest communities (i.e., boreal communities) are affected by environmental heterogeneity; (2) the spatial scale at which we estimated environmental heterogeneity (i.e., watershed) was not large enough to include streams that differ in a way that would matters for community assembly; (3) we have not measured the appropriate environmental variables – although we think we did; and (4) we did not investigate this issue with the appropriate tools.

Regarding this last issue, we think a good strategy would be to measure the strength of the relationship between community composition and raw environmental variables within each watershed (e.g., a R^2 values from an RDA) and then regress those values against community size. Following our expectations, one should find a positive relationship between community-environment strength (R^2 values) and community size. Unfortunately, although we had thought about this, we have not run this analysis because our sample size was too small (N = 5 stream per watershed) to get good estimates of R^2.

In the fourth paragraph of their discussion, the author discussed the effect of high dispersal rate as a potential drivers explaining higher similarity in species composition among communities than expected by chance. I fully agree with the discussion but why authors did not find any effect of the "spatial extent" variable (ie the distance average among communities within a watershed)?

R. We think high dispersal rates occur within all watersheds, not only in some specific watersheds. To find a significant effect of the "spatial extent" variable, dispersal rates would need to be different among watersheds. We don't think this is the case. In addition, watersheds are too small to impose serious dispersal limitation and do not differ too much in spatial extent. In Brazil, maximum distances between pairs of streams within watersheds varied from 2.48 to 8.86 Km, whereas in Finland it varied from 12.77 to 109.5 Km.

Finally, in introduction and summary authors talked about interaction among factors (such as ecological drift and niche selection for instance) driving species assemblage. But this interesting idea is quickly abandoned in the rest of the article. While it's not presented directly like this, my view of the discussion is that interaction between niche selection, dispersal rate and ecological drift determine species assemblage in the study. It is likely that community size modulates the effect of niche selection and dispersal rate on species assemblage (as the authors suggested in some parts, eg "community size ... may mediate the interplay between deterministic niche selection and ecological drift as drivers of beta diversity in tropical and boreal metacommunities"). Authors can and should test for that directly by adding interacting effect between community size and environmental heterogeneity as well as between spatial extent and community size. Considering the current interpretation of the authors, we can expect that the effect of spatial heterogeneity and spatial extent increase with community size. It will demonstrate that large community size are more determined by deterministic processes. Be aware that environmental heterogeneity captured the abiotic dimension of the niche selection but not species interaction, as well as that distance used to compute the spatial extent should consider the kind of dispersal of the genera studied (fly, stream vector,...).

R. If we understood this, you suggested a model like this:

beta dev ~ size + env.het + spa.ext + size*env.het + size* spa.ext

Our sample size (n = 20) is simply too small for a model with 3 main effects and 2 interactions. So, we rephrased the Introduction section to not create this expectation (interactions among processes) on readers' minds.

2) Controlling or correcting for sampling bias in beta diversity metrics:
The authors compared species assemblages among different streams, watersheds and regions which are characterized by different habitat characteristics. As a consequence, depth, width and velocity of streams are likely different. The authors have corrected one of their community metrics by the stream width to account for habitat size. It's a good point to do that because larger is the sampled habitat size higher is the probability to catch individuals and new species (or genera in their case). But stream width is only one dimension of the habitat size. May be it is more important to account for both stream velocity, depth, and width in order to correct metrics by the volume of water filtered during sampling (more they filter water higher is the probability to observe individuals and species; except if high stream velocity is considered has an environmental pressure which constrain species diversity and abundance). Difference in water volume filtered among streams could artificially increase the beta diversity of watershed as well as the difference of beta diversity among watersheds, and as a consequence bias the results. One solution could be to correct metrics but in this case why authors have corrected the size community metric only while beta diversity is likely impacted by this potential bias too? Another solution could be to not correct metrics. For example, they could use linear mixed effect model with a qualitative variable in random effect controlling for habitat size.

R. Thanks for your comment, but the sampling procedure we adopt was not directly influenced by flow or volume. The sampling procedure is as it follows (L. 178):

"At each of the 100 stream sites in both regions, we took a 2-minute kick-net sample (net mesh size: 0.5 mm), which was composed of four 30-seconds sample units obtained in the main microhabitats at a riffle site (e.g., which considered differences in current velocity, depth, benthic particle size and macrophyte cover). The four sample units were pooled, preserved in alcohol in the field and taken to the laboratory for further processing and identification".

So, basically, we sampled insect larvae that live attach to rocky substrates like pebbles, cobble, and gravel using standardized effort (time and area). Thus, it is very unlikely that our standardized sampling effort would bias our estimates of beta diversity. But more important, all measures of community size yielded very similar results.

Minor comments:
INTRODUCTION:
-"A solution is to use a null model to produce expected values, contrast observed and expected values and use the difference between them as estimates of beta diversity; called beta deviations hereafter (Kraft et al. 2011, Myers et al. 2013, 2015, Catano et al. 2017). In this case, positive and negative values of beta deviation indicate that communities are more dissimilar and less dissimilar than expected by chance, respectively. Beta deviation values close to zero indicate communities are as dissimilar as expected by chance (Kraft et al. 2011, Chase et al. 2011, Catano et al. 2017, Petsch et al. 2017)."
Here, authors provide only a solution for the first issue described (sampling bias) but not the second one (beta diversity index accounting for both compositional and abundance changes).

R. You are right. We included the following sentence to solve this (L. 108).

"A solution to the second issue is to analyze the data with dissimilarity coefficients that take into account species composition and species relative abundance (e.g., Bray-Curtis; Anderson et al. 2011). Such dissimilarity coefficients provide complementary information regarding the main mechanisms responsible for β-diversity patterns (e.g., Siqueira et al. 2015)."

- may be at the end of the introduction authors could briefly tell that they study the issue through a modeling approach testing for community size, environmental heterogeneity and spatial extent. I mean we know that authors consider these two last variables since methods only while it's an important good point for the present study. Added that in the summary could be good too.

R. We included a sentence about that at the end of the introduction.

METHODS
-authors should provide maps in SI in case of readers have no access to Heino et al. 2018 (like me).
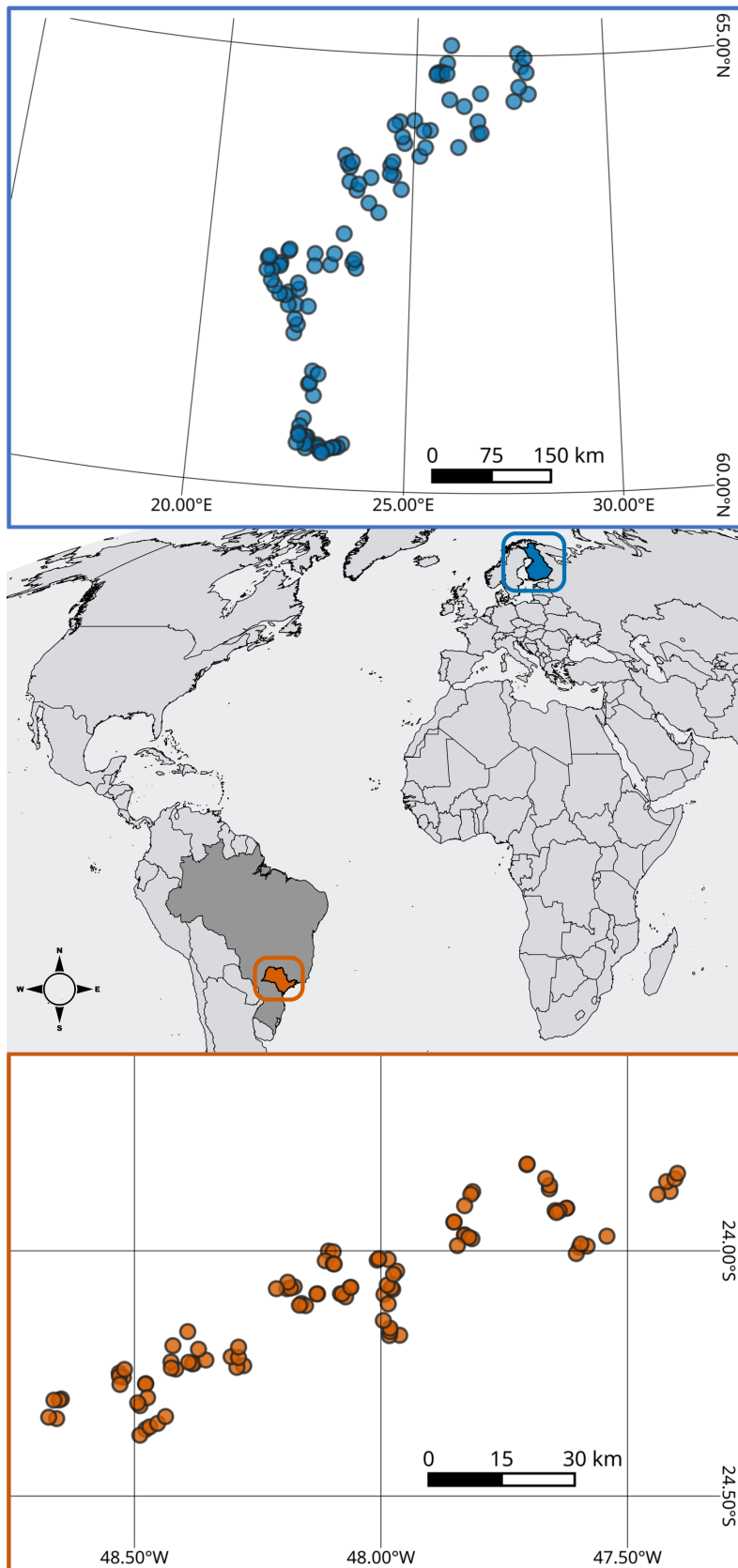
R. Done. We added a map as figure S2.

Figure S2. Geographical location and spatial distribution of streams in São Paulo State, Brazil (circles in vermilion) and Finland (circles in blue).

-Regions have the same areas? And watersheds? I mean if there are high differences in areas it could impact beta diversity among watersheds as well as the species pool, and as a consequence the results.

R. The sampled regions do not have the same area. In Finland, sites were located along a region of ca. 500 km and 300 km in north-south and in east-west directions, respectively, while in Brazil, they were distributed along a region of ca. 70 km in north-south and 120 km in east-west directions, respectively (Heino et al. 2018). Watersheds also differ in spatial extent between regions, being smaller in Brazil. In Brazil, maximum distances between pairs of streams within watersheds varied from 2.48 to 8.86 Km, whereas in Finland it varied from 12.77 to 109.5 Km. However, we don't think this is a major issue in this study as our analyses were done separately for each region. In addition, our null models yielded very similar results when we defined the species pool as all genera occurring in the whole region or when the species pool was defined as each watershed. Please, also note that we included spatial extent in our models. Thus, variation in watershed areas, within each region, was taken into account.

-"Also, we estimated the median population size per stream, average it within watersheds, and defined it as another measure of community size." What is population here? The number of individuals belonging to a genus?

R. It is the number of individuals per genus. We noted that recommender Eric Harvey and the other reviewer also thought this and other definitions of community size were confusing. Because results were very similar no matter the definition we adopted, we decided to remove theses sentences from the main text.

-"Fitted models provided similar results with all measures of community size and, thus, we show here results based on the former measure." So authors presented results based on the community size corrected by habitat size?

R. No, we presented the results based on community size defined as the mean number of individuals across five streams within a watershed. We added this information to the main text (L. 189).

-incidence-based beta deviations: "beta-deviations were calculated as an index rescaled to range between -1 and 1".OK but how did authors compute the beta deviation concretely?

R. Sorry, this was indeed unclear. Now the text reads like this (L. 202):

**"*To estimate incidence-based β-deviations that accounted for random sampling effects, we used a modified version of the Raup-Crick coefficient following Chase et al. (2011) and the procedures described in the package vegan (Oksanen et al. 2018): (i) we defined the genus pool as all genera occurring in each region and their observed occupancy across all 100 streams; (ii) an algorithm estimated the number of genera that any pair of streams share (SGobs); (iii) an algorithm assembled local stream communities by randomly sampling genus from the pool until reaching the local (observed) genus richness and by using the observed genus' occupancy frequency to determine the probability to sample a*

genus; (iv) step (iii) was repeated 10,000 times to generate 10,000 random matrices of genus composition and, posteriorly, the number of shared genera between each pair of streams within each watershed (SGexp); (v) β-deviations were calculated as the number of random draws in which SGexp ≥ SGobs divided by the total number of random draws. This index was rescaled to range between -1 and 1, "indicating whether local communities are more dissimilar (approaching 1), as dissimilar (approaching 0), or less dissimilar (approaching -1), than expected by random chance" within each watershed (Chase et al. 2011). As with β-diversity, this procedure resulted in 20 values of β-deviation per region, with one value per watershed (Fig. S1)."

-abundance-based beta deviations: "the (local) total abundance": is it the community size, ie the total number of individuals?

R. No, community size is the mean number of individuals across five streams within a watershed.

-why accounting for species frequency occupancy in incidence-based beta deviations but not for abundance-based beta deviations when computing null model?

R. Sorry we were not clear here. The algorithm randomly assigns individuals to local communities until reaching the observed community size, while preserving the overall species-abundance distribution in the region and the total number species. We tried to make the text clearer in this respect.

-"Results were similar with both definitions of species pools, and thus we only show results based on the former definition." So authors presented results based on region?

R. We removed this sentence because we don't think it makes sense to define the species pool as the watershed. They were too small.

-DISCUSSION:
-"probably together with deterministic assembly processes, as beta deviation values were different from zero"
Difficult to really judge because authors did not test for that. May be they could look at the significance of the intercept. If it is different to 0, it likely means that community diversity differences are more determined by non-random processes even in case of smallest communities. If it is not the case, so random processes drive differences between genus diversity among communities.

R. Here we are specifically referring to beta deviation values being higher than zero. As we explained in the methods section, according to Chase et al. (2011), only values close to zero would indicate that random processes drive composition dissimilarity. But as the text was confusing, we decided to remove that part of the sentence, which now reads as (L. 326): "*As findings generated by null models are our best approximation of patterns generated by stochastic processes (Kraft et al. 2011, Chase et al. 2011), we suggest that demographic stochasticity plays a major role in small ecological communities (Orrock and Fletcher 2005, Orrock and Watling 2010, Gilbert and Levine 2017).*"

-"Also, in general, these results are in line with our predictions,"
expectations/assumptions sound better than predictions.

R. Done. But just to be clear, here "predictions" is not being used as a modeling exercise (which involves validation techniques with independent data); e.g., as if we were using a mathematical technique to predict new values of a variable in a future scenario. Here, we are using "predictions" to refer to what would be our results, given our hypothesis.

-"dissimilarity should be low when niche selection is spatially constant (e.g., harsh conditions within the metacommunity, Chase 2010)".
Authors refereed to a specific example with low environmental heterogeneity (as they wrote) but with extreme conditions which can strengthen low beta diversity. In this case it's more the high environmental pressure which explain low beta diversity than spatially homogeneous environmental conditions.

R. Maybe. One would need to investigate this, as it is actually an interesting subject. But if high environmental pressure is variable along a spatial gradient (i.e., selection is not constant), one could also find high beta diversity. We modified the text, as Chase 2010 does not affirm conditions were really harsh (L. 372).

"*First, dissimilarity should be low when niche selection is spatially constant, as the environment maximizes the fitness of a few species (Vellend 2016).*"

-"Flenner and Sahln (2008) estimated annual rage expansions of up to 88 km" => range

R. Done.

-"If these inferior competitors have high dispersal rates, a trade-off suggested by theoretical models(Cadotte et al. 2006), then they would have a higher chance to occur in some small communities within the metacommunity." In this case authors refer to small communities where stochastic processes are predominant. So why talking about "high dispersal rates" while dispersal rate is by definition non constraining for species assemblage in case of stochasticity?

R. Our point here was that, according to theory, inferior competitors usually have higher dispersal abilities. Better dispersers should be able to (at least) colonize more sites within a metacommunity. However, because they are inferior competitors, if niche selection is the main force driving assembly, they would not be able to maintain populations in a number of sites within the metacommunity. If demographic stochastic plays a major role, however, then these inferior competitors would be able to maintain populations, at least, in some sites – because they are good dispersers and because demographic stochasticity allows them to not be excluded by superior competitors.

However, after considering some comments in this review, we decided to not be so emphatic about competition. Now we are simply referring to changes in the "*the*

alteration of the occupancy frequency and relative abundance of species with different fitness" (L. 331). Hope that clarifies your point.

Thanks again, Romain. Your comments were well appreciated and made us think about a number of important issues. Hope you are satisfied with our responses.

The following notes were made by reviewer Dr. Kevin Cazelles.

Siqueira et al. investigate the relationship between beta-diversity and community size (total number of individuals) for aquatic insect communities of two regions: Brazil and Finland (note that the data set is described in Heino et al. 2018). The main findings are:

1. a negative correlation between beta diversity metrics and community size for Brazilian communities but not for Finnish ones;

2. incidence-based beta deviation (i.e. a score based on the difference between observed beta diversity and the expectation under a null model) is negatively correlated to community size only in Brazil but abundance-based beta deviation (Bray Curtis index) is positively related to community size in both regions.

Based on these results, the authors conclude that the smaller the community size the bigger the role played by ecological drift.

I found the manuscript overall clear, the authors have used high scientific standards and they provide data and code to reproduce their analysis. That being said, I think the analysis performed weakly support the conclusions. I also think the manuscrit is missing important pieces of information. For these reasons, I believe that the current version of the manuscrit is not suitable for a recommandation by PCI.

R. Below, we provide responses to your comments, hoping we are able to change your view about our manuscript. In general, we think that most of the comments were addressable (even the one about the weakness of our results to support our conclusions, which we do not fully agree by the reasons described below).

1 Major comments
1.1 Ecological drift

I am not convinced that the analysis support the conclusion that ecological drift is higher in smaller communities. As far as I understand it, the analysis (as described in Chase et al. 2011; Kraft et al. 2011) do not allow the authors to determine the nature of the mechanism behind the deviation observed. I agree that ecological drift is one plausible explanation but there are alternative explanations that are as much convincing as the ecological drift. For instance, the variations in dispersal capacities of species and the connectivity within watersheds could explain well the results obtained. This is discussed in one paragraph page 16, but it could also have been an option to frame the paper. My point is that the authors

have shown a relationship between beta deviation and community size, this is an interesting result but they cannot conclude that it is due to ecological drift. It may be ecological drift, it may also be dispersal or it may be because of the topology of the whole ecological network, or a mixture of these mechanism. My opinion is that the authors should acknowledge this and should not neglect alternative hypotheses.

R. Indeed, by only looking at beta deviation patterns, as an exploratory exercise, one cannot be sure about the underlying mechanisms. This cannot be done in any observational study relying in individual evidence. But we did not simply do this. We derived specific hypotheses based on theory, mechanistic models and previous empirical evidence, and went to the field and sampled data to test these hypotheses, trying to isolate confounding factors via using identical methods in two regions/countries. In addition, we used different evidence, from different analyses to support our inferences. One main hypothesis in our study was that smaller communities are more variable in species composition due to the effects of demographic stochasticity. To test that, we used standardized sampling effort in 100 streams (same area and time effort in each stream) in two different countries. Within each country, watersheds had similar dendritic structure and stream size, and watershed and region extent were small. That is, community size does not coincide with within watershed connectivity or the topology of the network – we thought about having replicates of watersheds before going to the field.

Also, we are not sure how dispersal capacities of species would make only smaller communities more variable in species composition than larger communities, as you suggested. One possibility would be that dispersal limitation (or dispersal surplus) plays a major role only in smaller communities, weakening selection. But in both cases, limitation or surplus, dispersal would produce spatial patterns that would be different from what would be expected by a niche deterministic model. That would likely result in (1) beta deviation being lower than zero, suggesting communities are less dissimilar than random expectations (as we observed in larger communities), with dispersal surplus being a strong candidate driver; or (2) beta deviation being higher than zero, suggesting communities are more dissimilar than random expectations (as we also observed in larger communities), with dispersal limitation being a strong candidate driver. We don't think dispersal limitation could be responsible for scenario (2) in our study, as watersheds were not large enough to allow strong dispersal limitation. In Brazil, maximum distances between pairs of streams within watersheds varied from 2.48 to 8.86 Km, whereas in Finland it varied from 12.77 to 109.5 Km (this information was added to the manuscript). There is only one scenario where dispersal limitation would make communities as variable as random expectations – a neutral scenario, where the environment does not play an important role, which, to some extent, is in line with one should expect when ecological drift is a major structuring process.

So, in this sense, we respectfully disagree with you. We think stochasticity in demography was responsible for making small communities more variable and closer to random expectations. We also think dispersal surplus was responsible for making incidence-based beta deviation of medium to large communities negative – i.e., genus composition of these communities was less similar than null expectations. Finally, we think niche selection was responsible for making abundance-based beta deviation of medium to large communities more dissimilar than expected (positive values).

1.2 Information missing / major lack of clarity
1.2.1 Information about the aquatic insect communities

I understand that there is a published study that describes the data set (Heino et al. 2018). That said, from a reader perspective I think a little bit of information about the communities is required. It could be something like figure S5 in Heino et al. (2018) that would give a sense of the spatial turn over of the communities. So far there are only two figures, so I think there is room for two or three more.

R. We added information about spatial extent: "*In Brazil, we sampled 100 streams distributed among 20 watersheds located in the southeastern region of the country – i.e., five streams per watershed, with spatial extents of 70 km in north-south and 120 km in east-west directions, respectively.*" (L. 151) … "*The study sites in Finland were situated in the western part of the country. We sampled 100 streams that were distributed among 20 watersheds (as described above) with spatial extents of ca. 500 km and 300 km in north-south and in east-west directions, respectively*" (L. 165).

About streams (L. 171): "*Streams were generally of the same order within each watershed, but varied a bit among watersheds, including 2nd and 3rd order streams in Brazil, and a few 4th order streams in Finland. In Brazil, maximum distances between pairs of streams within watersheds varied from 2.48 to 8.86 Km, whereas in Finland it varied from 12.77 to 109.5 Km. Most in-stream abiotic variables varied within a similar range and had similar mean values between regions, except conductivity, total nitrogen and total phosphorus, which were much higher in Finland*".

And about communities (L. 184): "*We sampled 16,113 individuals, distributed among 83 genera in Brazil, and 86,048 individuals (77 genera) in Finland. The mean number of genera per stream was 17.84 (standard deviation = 7.46) in Brazil and 14.01 (sd = 5.07) in Finland, while the mean number of individuals per stream was 181.50 (sd = 111.38) and 886.57 (sd = 700.73), respectively (Heino et al. 2018).*"

We are also including a map now (see above).

1.2.2 Community size and stream width

Page 7, I read:

"Because streams within and among regions differ in width and this could be viewed as measure of habitat size, we multiplied local community size by stream width, averaged it within watersheds, and defined it as an alternative measure of community size."

I do not understand why the community size was multiplied by stream width. I think that in order to correct this potential bias, one should divide rather than multiply. But more importantly, I do not understand the need for recasting the definition of community size. It would be better to use community size as defined in Vellend (2010), perform the analysis and then check if there is in fact an effect of stream width.

R. Apparently, this was causing confusion – the other reviewer and recommender Eric Harvey also made similar comments. So, as results were very similar no matter the definition of community size, we decided to remove the text about alternative definitions of community size and used the original definition by Orrock and Watling (2010). Now the text reads (L. 189): "*We adopted the definition of community size provided by Orrock and Watling (2010) and estimated local community size as the number of individuals sampled in a stream site. However, as β-diversity was estimated at the watershed scale (i.e., considering 5 stream sites; see below), we averaged the number of individuals across five streams within each watershed. This resulted in 20 values of community size, one per watershed (Fig. S1).*"

1.2.3 Simple correlations

I would recommend to add simple correlations in the manuscript:

1. I recommend to investigate the role of average species richness per watershed. This is important given that beta-diversity metrics actually depends on the number of species.

2. I was surprise to see that there is a strong relationship between the number of species per watershed and the local community size in Brazil but not in Finland (see Figure 1). Can this explain part of the results? I think this relationship should be included in the manuscript and discussed.

3. The authors showed that there is no effect of site heterogeneity on the beta deviation but I think they should rather investigate whether there is a role of heterogeneity on the raw values of beta diversity.

R. Regarding your point #1, indeed some beta-diversity metrics are affected by species richness and this is another reason for using Raup-Crick beta deviation that accounts for random sampling from the species pool. This metric controls for this exact problem by simulating beta diversity values while holding local richness and the number of shared species constant (Chase et al. 2011), but sampling from the whole species pool. This property makes this metric appropriate to compare beta deviations from sites with different alpha and gamma diversity.

Regarding your point #2, originally, we thought about including this as another expectation in our study, as smaller numbers of species should be found in smaller communities (Vellend 2016). We decided to leave it out because the focus was on beta diversity. We must emphasize, however, that variation in species richness along the community size gradient in Brazil should not be the reason for the results we observed. The whole approach using null models eliminates any due to sampling effects on alpha diversity. Following your recommendation, we investigated the correlation between mean genus richness per watershed and community size and found a positive correlation in Brazil (Peason's r = 0.81; p < 0.001), but not in Finland (Peason's r = 0.42; p = 0.06). These results reinforce our inferences about the role of demographic stochasticity in smaller communities only in Brazil.

Regarding you point #3, thanks for this suggestion, but our focus is on beta deviation as we see it as a proper metric to estimate beta diversity without bias due to random sampling effects on the species pool. Although we decide to not include this in the manuscript, the following were the results considering the relationship between beta diversity (incidence-based) and environmental heterogeneity:
In Finland (b = 0.53; t = 2.7; p = 0.015);
In Brazil (no relationship).

1.2.4 Algorithms

Regarding the null model introduces in Chase et al. (2011), the authors wrote: "(i) we defined the species pool as all species occurring in each region;"

But according to Chase et al. (2011)
"At the same time, it is not advisable to use a regional species pool that is so large (e.g., all of the species of a particular group across biogeographic zones) that all communities would have exceptionally low βRC values."

How do the authors check that they are not using a pool of species that is too large?
Also, I do not fully understand the procedure described in Kraft et al. (2011). Do the authors pool all species or all individuals? Do the abundance of a specific species have an influence on the probability of drawing it (like in a neutral procedure)?

R. Good point. For sure, we are not dealing with a biogeographic zone. In Finland, sites were located along a region of ca. 500 km and 300 km in north-south and in east-west directions, respectively, while in Brazil, they were distributed along a region of ca. 70 km in north-south and 120 km in east-west directions, respectively (Heino et al. 2018). Also, using the watershed as the regional species pool would not make sense here, as they were too small and only included 5 streams. This would lead to a very reduced gamma diversity, from where the null model would sample. As watersheds had different gamma diversity, we thought it would be better in this case (Raup-Crick) to use a unique species pool – i.e., we did not want different watershed gamma diversity to affect our estimation of watershed beta diversity.

Regarding abundance-based beta deviation, yes, all individuals were pooled and genus relative abundance in the entire region are preserved during the null assembly. Sorry we were not clear in our original description.
The text now reads (L. 202) : "*To estimate incidence-based β-deviations that accounted for random sampling effects, we used a modified version of the Raup-Crick coefficient following Chase et al. (2011) and the procedures described in the package vegan (Oksanen et al. 2018): (i) we defined the genus pool as all genera occurring in each region and their observed occupancy across all 100 streams; (ii) an algorithm estimated the number of genera that any pair of streams share (SGobs); (iii) an algorithm assembled local stream communities by randomly sampling genus from the pool until reaching the local (observed) genus richness and by using the observed genus' occupancy frequency to determine the probability to sample a genus; (iv) step (iii) was repeated 10,000 times to generate 10,000 random matrices of genus composition and, posteriorly, the number of*

shared genera between each pair of streams within each watershed (SGexp); (v) β-deviations were calculated as the number of random draws in which SGexp ≥ SGobs divided by the total number of random draws. This index was rescaled to range between -1 and 1, "indicating whether local communities are more dissimilar (approaching 1), as dissimilar (approaching 0), or less dissimilar (approaching -1), than expected by random chance" within each watershed (Chase et al. 2011). As with β-diversity, this procedure resulted in 20 values of β-deviation per region, with one value per watershed (Fig. S1)."

2 Other Comments

• p.2:
"However, although beta diversity and community size were strongly related in both regions, the type of relationship varied according to the type of dissimilarity coefficient."

I think this is slightly confusing because the authors are actually referring to results for the beta deviation.

R. We changed the text to (L. 32): ***"Null models allowed us to estimate the magnitude to which beta diversity deviates from the expectation under a random assembly process while taking differences in species richness and relative abundance into account, i.e., beta deviation. Beta diversity of small communities was consistently higher but closer to null expectations than β-diversity of large communities. However, although β-deviation and community size were strongly related in both regions, the direction of the relationship varied according to β-diversity metrics"***

• p.4
"Thus, to properly analyze the relationship between beta diversity and community size, we need estimates of beta diversity that account for differences both in species richness and species relative abundance. A solution is to use a null model to produce expected values, contrast observed and expected values and use the difference between them as estimates of beta diversity."

I am confused by the sentence as it sounds like beta deviations are better metrics of beta diversity whereas there a metric that compare observation (classical indices of beta diversity) to a null models. I guess my confusion is due to the wording of these sentences.

R. Yes, we are stating beta deviation is better than the raw one for inferences of process that may drive variation in community structure. The text in the introduction coarsely describes the procedure to obtain beta deviation. The full procedure is described in methods section.

• p.8:
"We tested whether beta deviation was related with community size by using ordinary-least-squares regression models."
Why are you using "ordinary-least-squares regression models" instead of "linear model? (I know that here they are equivalent but you are actually using lm in your code not nls)

R. Right, we changed to "linear regression models".

• Figures: I think it would make sense to pool observations for Brazil and Finland on the same plots (like in figure 1 I've created). Hence the reader would easily see the full gradient of community size. Also I think it would be better to add horizontal and vertical errors bars to give a sense of the variance along both axis for each watershed.

R. Thanks. Now figures include both regions in the same panel.
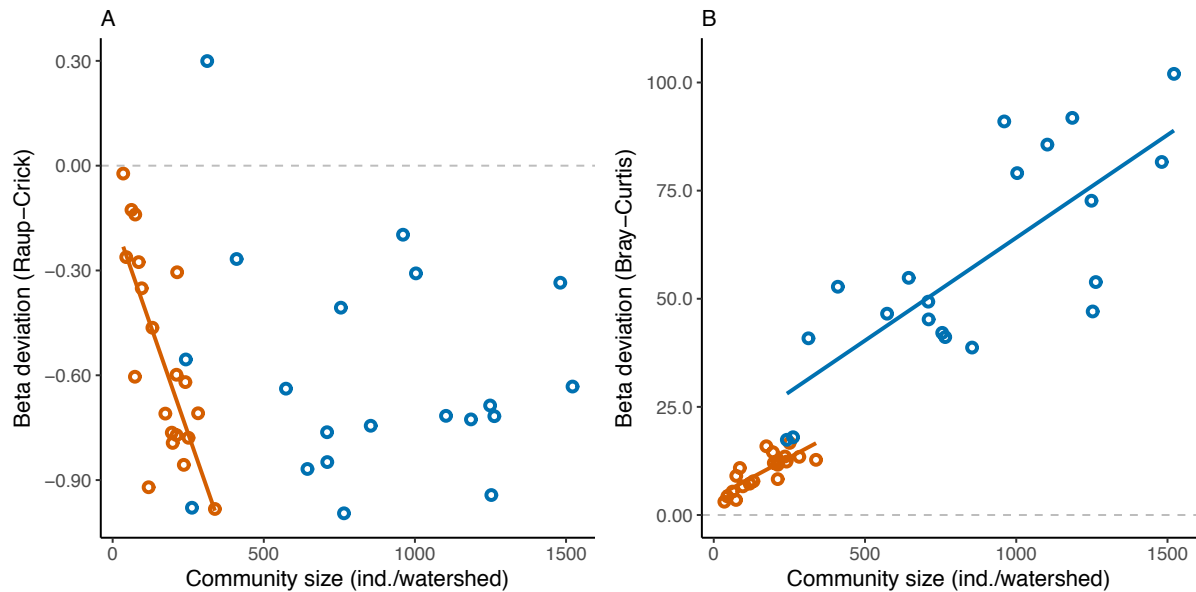

Figure 3. (A) Incidence-based (Raup-Crick) and (B) abundance-based (Bray-Curtis) beta deviation-community size (average number of individuals per watershed) relationships within tropical (vermilion) and boreal (blue) stream watersheds (n = 20 for each region). The average number of individuals per watershed was calculated with a sample size of 5 streams. The dashed grey line indicates expected beta diversity under null assembly.

Regarding the error bars, we also tried and think they look worse and do not add essential information that would change the main interpretation. But, again, we are open to discuss this.
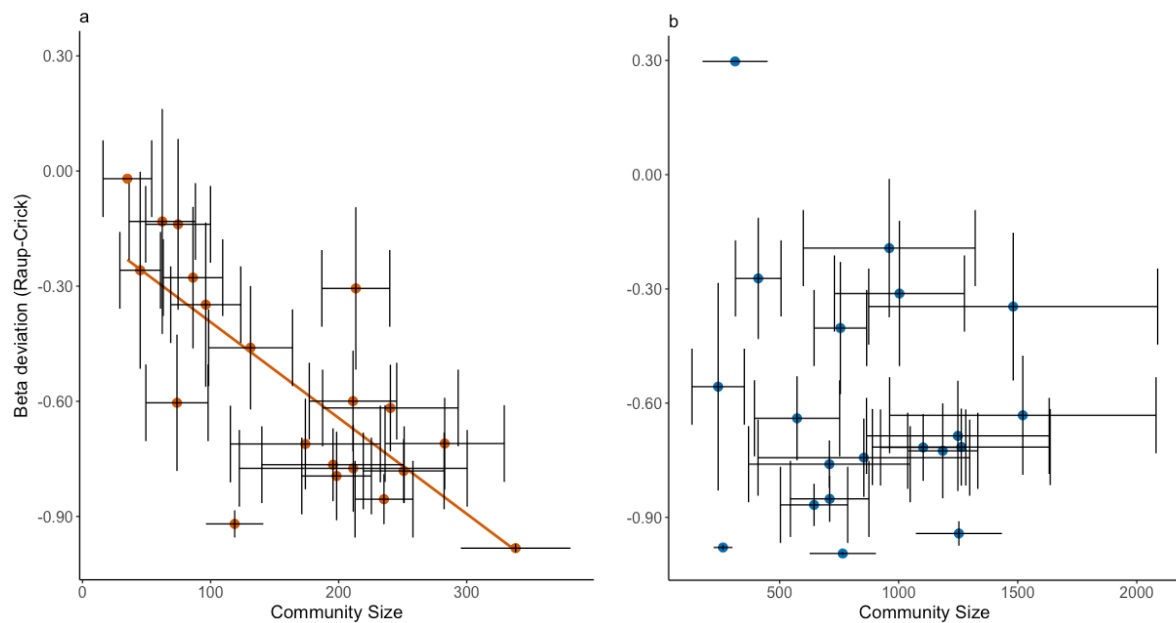
• Data and code:
I very much value the efforts that have been put into making the analysis reproducible, that is great. I have a few suggestions:

• In the R scripts, the authors should provide the description of all function arguments. Currently, it is hard to follow the steps just because of this;

R. Arguments of helper functions are now better described in the file helper_functions.R. All other functions we used have their arguments described in their help functions, within each package where they belong.

• the link to the FigShare repository
https://zenodo.figshare.com/articles/Community_size_affects_the_signals_of_niche_selection_ should be added;

R. The doi link to the Zenodo repository was added.

• I would add a brief description of the files in Metadata_SacleBio.odt;

R. Done.

• also, there is one typo ("assinged") in catanos_betadev.R and 2 ("communuty" and "threatments") in Script_beta_diversity_deviation_Bra_Finn.R.

R. We have corrected that, except "communuty" because we did not find it.

Thanks again, Kevin. It was a detailed review with important points that certainly improved the quality of the manuscript. Hope you are satisfied with our responses.