**Reviewer #1: I was excited to read the MS by Thomas and colleagues, which offers novel insights and scales of consideration for fungal foliar and wood endophyte communities.  The most novel contribution of the paper, in my opinion, is the detection of landscape-scale variation of community dispersion.  This is a system and topic of interest to Fungal Ecology's readership.**  
  
**I have a few suggestions for how the paper might be edited for clarity, and a couple of statistical points that the authors may wish to consider in a revision: mostly the potentially confounding factors of topography, distance and host identity. I also feel that the authors spent much of the paper reporting their negative results from the whole community analyses.  Negative results are fine, but with 13 figures, I suggest condensing these to the most critical and moving most to supplements.**

Removed.  
  
**I was unable to access the source code for this paper, and I did not see where/if the data were deposited to the SRA or another sequence repository.**

The information for the read libraries was supplied in the data statement, the link to which is the final page in the submission pdf. The reads are deposited in two different repositories:

1) as suggested by Elsevier, the wood reads are stored with Mendelay data. However, wedouble checked, and found a typo in the wood reads doi. The doi listed in the data statement for the wood endophyte reads is <<http://dx.doi.org/10.17632/czvrfcj377.3>>. This is **incorrect**. The actual doi for the wood endophyte reads is:

<<http://dx.doi.org/10.17632/czvrfcj377.2>>

This will be corrected.

2) Leaf reads are held by figshare, listed correctly at the time in the data statement. The doi again is:

<https://doi.org/10.6084/m9.figshare.3208252.v1>

**In the future please submit your paper with line numbers to facilitate reviews.**

From the author instructions, <<https://www.elsevier.com/journals/fungal-ecology/1754-5048/guide-for-authors#4001>> which at the time of initial submission said the following:

*“Manuscripts should be double-spaced throughout. Do not include line numbering in the manuscript.”*

**First paragraph is very long.  Could use a topic sentence.**

Paragraph has been shortened. The first sentence is intended to be a topic sentence.

**Field Methods: So host tree was not explicitly considered in the sampling scheme?**

Host was not explicitly considered for both statistical and logistical concerns; it was entirely unknown before sampling. We undertook too large a study in too biodiverse a site to sample in a way that could balance hosts evenly. When ranked by abundance, tree species at the Fusan FDP follow a typical rank-abundance curve, with a small number of highly abundant tree species and a large number of relatively rare trees. Some hosts exist only in small clusters on the landscape, or are confined to unusual habitat. Attempting to balance our host numbers beyond a few abundant tree species would have resulted in an extremely contrived, clustered sampling pattern, aggravating the concerns about bias due to autocorrelation in the microbial communities.

We also needed a sufficient diversity of hosts to make our co-occurrence analyses robust. We needed enough other hosts, even if smaller numbers, to tell if a microbe was tending to co-occur with the hosts for which we had sufficient sample sizes to spatial analyses. The numerous other hosts with small sample sizes picked up in our study allowed us some extra statistical confidence to say that the core microbes we detected in *Helicia* were indeed tending to co-occur with *Helicia* more often than they seemed to be in other hosts.

**Are the hosts spatially autocorrelated?**

As I’m sure Reviewer #1 is aware, the term autocorrelation is sufficiently broad to require further explanation of this question. Vegetative community does show strong spatial patterning, as is easily seen in fig. 1 but not quantitatively analyzed in our manuscript. A quick review of the individual tree species distributions on the plot also reveals obvious spatial patterns at several scales. Much of this patterning is discernible by the eye to be due to external ecological condition, and some excellent analyses by Su et al. support this2, creating “spatial dependence” in the tree distributions. Some of the patterning could be due to “autogenic” or “internal” autocorrelation.

We would like to know a bit more about which type or scale of autocorrelation in which Reviewer #1 is most interested, but wecan imagine a they would be interested a classical univariate spatial analysis of individual tree species (Moran’s weof tree density, Nearest-Neighbor, kernel-density-estimate or Kriging-type analysis), or a multivariate, tree community approach using tools similar to the spatial analyses in used in this paper (PCNMs and mantel tests of community dissimilarity). Further spatial analysis of spatial patterns of trees at Fusan would be a complex but very worthwhile project. It is unfortunately beyond the scope of this paper, and webelieve the data from the tree census at this time are still under the supervision of Su et al. who are working on various projects with this data, though access to this data may be available upon request.

webelieve weunderstand Reviewer 1’s reasons for asking this question, however. It is important to ask whether any spatial patterns observed in the fungi are results of induced spatial dependence on their tree hosts and the habitat that these trees prefer,. The result of this would be lots of error due to autocorrelation, potentially causing false positives in our ecological analyses that are not due to the fungi but simply where the hosts tend to aggregate. To address this question we focused most of our analyses on the fungi of just one host.

Any remaining error due to autocorrelation that may have resulted from a clustered sampling scheme is then quantifiable in the spatial analyses. It is quantified in our study with the Mantel tests and the smaller scale PCNMs. In the case of multivariate Mantel tests, if assumptions are met, internal autocorrelation is often considered to be represented by the range of comparisons that are above the x-axis in the Mantel correlograms. In the case of PCNM/dbMEM analyses, the eigenvectors with smallest wavelengths are considered strong candidates for describing internal or autogenic autocorrelation, especially if they are not correlated with environmental data (Borcard et al 2011). Other remaining eigenvectors are orthogonal to these, and can considered for the purposes of ecological analysis independent of internal autocorrelation. Thus if we restrict our ecological observations to patterns that are larger in scale than these eigenvectors, we hope that we are not simply describing patterns due to internal autocorrelation. Both our mantel tests and our PCNM/dbMEM analyses point to ecological patterns of interest in our fungal community at a scale of at least ~150-200 m, and mantel tests indicate that positive autocorrelation effects are nearly non-existent by 150 m in both plants and wood.

**Bioinformatics:  Please explain what variance stabilization is.**

weagree with Reviewer 1 that this is a complex topic that many readers may want to know more about. However, it also now a fairly common term in microbial ecology when dealing with second-generation sequencer technologies. We’re not sure that we should take the extensive space within the manuscript to explain, especially given that it is generally suggested by the reviewers that weshorten this manuscript. A brief explanation has been added, and citations are given for readers who want to know more.   
  
**In the Bioinformatics section it's stated that the OTUs were converted to presence/absence.  But then it's stated that data used BC distance index and were Hellinger transformed… Can you please clarify which data were used for which analyses**

All read abundances of an observation of a fungal OTU in a sample were simplified to presence/absence, before any other analyses were performed. So analyses that involved the community matrices of endophytes – which is ***all*** of the analyses reported in this manuscript - were performed using presence/absence type community matrices, not with read abundances. This was due to our finding that differences in abundances of reads did not represent “actual” abundances in our positive control (see discussion here: < <https://doi.org/10.1101/184960> >). wehave added clarification in the text.

BC dissimilarity was used wherever a dissimilarity measure was required for a statistical test or for visualizations. Clarification has been added.

Hellinger transformations of community data were applied only when required by the assumption of a statistical test. This was only when applying redundancy analysis (RDA) to our community matrices indicated in the text, and not elsewhere.

These transformations are not mutually exclusive, obviously. In several tests all three were applied to the data. More clarification has been added in-text, hopefully it helps.

**-Trends in numbers of observations parallel patterns in OTU diversity (Fig. 4); if a class of fungi contained a large diversity of OTUs, it also tended to be observed often throughout the study site.  
  
This sentence is awkward and nothing about "occupancy" is shown in figure 4.**

The sentence was removed.

**-For all of the PERmanova analyses, this is a measure of both dispersion and "location" so comparing categories with different sample sizes will almost always lead to significant differences because higher N samples will have greater dispersion.  I suggest using a type III sum of squares (package(cars) can handle this), or check your dispersion using betadisper in vegan.**

We are aware of the possibility of confusing centroid-location and dispersion differences with PERMANOVA and other non-parametric, multivariate tests in the ANOVA family. The sample design is highly unbalanced (for reasons listed above), and dispersions were very different (see scripts, section “NMS/PERMANOVA Ordinations, Host”). As shown in the manuscript, all 2D NMS solutions estimate separate locations of centroids/dispersion in dissimilarity space among several hosts, with the most extreme example being *Cyathea*-associated leaf endophytes. In general, we believe that while effect size (R2) from the PERMANOVA model for host effects is probably exaggerated by differing dispersions, the p-value resulting from the pseudo-F comparisons for host-effects are not unrealistic, given the separation of groups shown in all NMS solutions.

In the initial PERMANOVAs are each single factor tests. In the case of host, host is the single factor, host species are levels. we did not nest them. we are uncertain how to apply a different type of sum of squares for the treatments or error in a “simple” ANOVA. My (admittedly poor) understanding is that alternate SS methods are generally applied when dealing with multifactorial experimental setups – is the reviewer asking that we treat sample size as a second factor in the ANOVA? we would appreciate further instructions on this. Even if we could follow this recommendation on a classical univariate ANOVA, we have not heard of alternative sum-of-squares methods developed for in the non-euclidean (Bray-Curtis) space of dissimilarity-based analyses such as PERMANOVA3. The car package has some support for alternate SS in MANOVA, but we can find no tools for PERMANOVA. The R car package is for general linear models, not for dissimilarity-based pseudo-linear models like PERMANOVA . We are not qualified to propose our own improvements to the PERMANOVA algorithms, or to make modifications to the existing code in the vegan package. This is perhaps something to bring up with Anderson, Oksanen, et al… So again, further instruction would be helpful.

Beta Dispersion was checked in the original scripts supplied with the initial manuscript, but the reviewer notes that they were not able to access to the supplementary material. In addition, even had they been available for Reviewer 1, the figures/results in these scripts were insufficient or absent in many places. This was due to some personal events that included a stolen computer and lost backups. we have made further efforts to reconstruct the original notebooks from the history of git commits. From this all original code was recovered, but some large gaps in figure outputs remain lost. As such, we rebuilt the scripts from ground up with figures. This was the main effort that took place during this round of revisions.

These rebuilt scripts include our process for observing multivariate dispersion in the microbe community dissimilarity matrix. Special detail is given to wood samples.

As with all microbiome studies, much of the art is in finding the signals in the massive noise. Here, much of the paper can be thought of as an exploration and denoising of this initial signal from host detected by the initial PERMANOVA/NMS tests. Taken alone, this signal of host-effects is, as reviewer #1 notes, not enough to write home about. Given the shortcomings of the PERMANOVA test, we dig further into the reasons for this signal, by decomposing the community matrix two into different categories of fungi, core and non-core. Plant microbiomes are incredibly, inordinately variable but when we do this binning, some order emerges. This subset of more “loyal” fungi that are identified through co-occurrence networks present themselves as the reason for the host-effects that are observed (shown in the BC comparisons in the last figure in the manuscript). They also show some interesting spatial patterning, unlike the chaotic microbial community at large.

**Better yet, add all of these components to your RDA model and calculate partitioned variance explained when all of these factors are considered together.**

A variation partitioning analysis was conducted for the environmental, host, and spatial (PCNM/dbMEM eigenvectors) matrices as explanatory variables for the fungal community matrix. This is, effectively, an RDA model (or rather several, compared). This was in the original manuscript, reviewed by reviewer 1. So further clarification from Reviewer 1 is needed for me to really answer this comment.

However, we should note here that we have decided to remove the language from the main manuscript around the variation partitioning, as it is mostly a negative (inconclusive) result, adding unnecessary complexity to the manuscript. Code and results for this analysis are still available in the first scripts.

In general the comments from reviewers have made me realize the PERMANOVA tests are over-emphasized in the manuscript – they are a minor part of the overall analysis. They were intended to show that a strong signal from host effects seems to be present, but that we must dig deeper into the spatial aspects of the sampling and behavior of “core” mycobiome to understand what may be behind this signal. As such, we have attempted to de-emphasize these test results.

**-It took me a couple of readings to understand the main point: that although community composition is not particularly spatially patterned, community dispersion (heterogeneity) is.  This very interesting and novel result, unfortunately, seems to be a bit buried in the long list and figures of negative or weak results when the entire dataset was considered.  I suggest highlighting this result more prominently and earlier in the paper.  Perhaps also consider reporting this in more straightforward language (like heterogeneity).  Couching this in terms of a "gradient" of "core" microbes requires some mental gymnastics.**

we believe that Reviewer #1 has missed the point that we’ve tried to make with this paper, or disagrees with it. If the former, I’ve tried to simplify and clarify with the revised manuscript presented. Unfortunately, we cannot discard the language and theory around the discussion of core microbiomes. The initial motivation and hypothesis for the study was the null hypothesis that core microbiomes do not exist in nature. Indeed, we began the study attempting to debunk the concept of a core microbiome and was then convinced that perhaps this is indeed a useful vocabulary for describing what we had observed. Though I’ve perhaps failed to convey this, the narrative is fairly simple.

I’ll summarize here and we hope that the revisions to the manuscript serve to also clarify the message. In addition, we have discarded other results, especially results that concern the importance of the hilltop, that may be distracting from this central narrative.

We sampled endophytic fungi from numerous host trees. We found:

1) Mycobiomes are incredibly chaotic and different, even within single hosts species and on such a small scale as the boundaries of the Fushan FDP.

2) Nevertheless, fungal communities are differentiated somewhat by host. To discern the reasons for this host effect, we focus all following analyses on the most numerous tree species, *Helicia formosana*.

3) We identify a very small subset of fungi, which we call “core” fungi, that co-occurr more often than expected by chance with their host, among thousands of other fungi that do not. This is the definition that Reviewer #1 notes requiring mental gymnastics, we hope we have clarified the language around this well enough so that it no longer does.

4) We explore the patterns of these core fungi. The initial signal of host effect noted above seems well explained by the presence of these fungi. There are also interesting differences in spatial patterns between wood and leaf “core”.

T**wo things you might want to address:  
  
1)    All of the spatial patterns in fungal communities appear to correlate with vegetation (composition, heterogeneity etc.)  Given the reasonably strong host effect found, is it possible that surrounding vegetation is shaping fungal community structure even when host is held constant? Could it be that "core" fungi aren't actually specific to host, but instead to an ecotype?**

Reviewer 1 may have misinterpreted some of our efforts and results concerning this. Vegetative community and our other composite environmental variable were not really useful in predicting differences in fungal communities. We hope that the new scripts clarify things a bit, language has been clarified in the manuscript itself, and here is a verbose explanation:

Some very small amount of variance in our wood and leaf endophyte communities was explained by our environmental variables when all tree host species were considered. Though statistically significant, the effect size of the correlations mentioned above are very weak in both our original and newer analysis, ranging from R^2 = 0.02 to 0.09, to the point that they might be considered a negative result. It is difficult to tell whether this is due to coarse resolution of our environmental data (20m x 20m grain), or whether these results are “real”, indicating that endophyte communities are not well predicted by environmental conditions at this spatial scale.

In addition, the environmental variables (vegetative and topographic), are highly correlated among themselves, and spatial patterns are not able to be taken into account with a test like a simple PERMANOVA. So we cannot easily interpret these small effects observed by the environmental PERMANOVA models. For this reason a partitioning of variance analysis was conducted, though it was vastly underpowered and we have removed it from the manuscript body as suggested. Code for it remains in the original scripts. Regardless, when we then focus on a single host, *Helicia formosana,* we lose all correlations between our environmental data and our fungal community.

We were unable to detect clear ecological drivers behind the endophyte community at large, but we are not surprised by this. We expect the microbiomes of trees to be complex in the mathematical sense, with many non-linear patterns, influenced at multiple scales in time and space, with chaotic behaviours in abundances and locations over time. There may simply be no order or manner of predicting patterns of most of the plant microbiome on this spatial scale and at a single point in time - virtually all spatial scales on the planet have been shown to be important to understanding microbiomes. Here we examine environmental predictors on a scale of ~700m or less, with most of our statistical power for detecting differences in microbial community lying in comparisons of communities at distances far less than this. However, the focus of this paper was on the detection and ecological description of possible “core” fungi, at the sampled scale. And for these fungi we uncover some spatial patterning that can be described relatively simply.

We did not supply an analysis of composition, in the sense of how different taxonomic groups vary among samples with distance and ecological conditions. Instead, our analysis were largely based upon community (dis)similarity. All community dissimilarity measures were conducted at approximately the species level, for which ITS is probably best suited. As with most environmental sequencing studies, very many of our OTUs were not identified with high confidence, which chokes our ability to conduct higher-than-genus-level taxonomic analyses, or even to do extremely robust comparisons based on species identifications. We do present a picture of the ratios of the classes present in fungi that have been identified (Fig. 3), but without accompanying ecological analyses. Higher level taxonomic analyses in environmental sequencing efforts are probably better done with a region of the LSU. This direction of inquiry also seemed unlikely to be interesting, given the low amounts of variation explained by environmental variables at the species-level.

**2)    Local patterns of diversity and richness will have a strong impact on community heterogeneity.  To what extent do your heterogeneity results reflect richness gradients?**

The reviewer requested more information on the process of variance stabilization above. A discussion of this is required to answer this question.

Environmental sampling using culture-free, direct-sequencing technologies such as were employed in this study are generally considered to be very incomplete sampling of microbiomes. To give an example, in the case of the leaves, three leaves were taken from the host tree of at each site, from among thousands of leaves present in the canopy. From each of these three leaves, a single 1 cm^2 piece of leaf lamina was removed, from the several hundred of square centimeters of leaf tissue present in these leaves. Column purification of DNA followed, representing a sampling event of the available fungal genetic material. ITS-region primers were then used to bind to a subset of the fungal DNA present during PCR, representing another sampling event. In the illumina flow cell, the lawn of sticky oligonucleotides binds to the ITS-region amplicons in a manner that is ideally random, representing yet another sampling of the population. Thus at several points this process is nested, with multiple sub-sampling events of the microbiome of the leaves occuring.

Any and probably all of the steps in this process introduce large bias into the relative numbers of fungal species amplified, and into the relative abundances among these fungal species within any given sample. Unfortunately, these errors due to sampling are not simple to model – some samples receive deep coverage, others do not, and their read “abundances” do not correlate necessarily with “actual” abundance, from whatever upstream step we can consider “actual” (DNA? PCR product?). In the same way, the number of species recovered from a sample does not scale linearly with its original diversity. Number of species recovered in a sample is more a function of depth of coverage (which as mentioned doesn’t correlate well with “actual” upstream abundance) and also varies directly with ITS copy number, ease of DNA extraction, etc. of each of the species of endophytic fungi.

Variance stabilization attempts to model some of these sources of bias and allow comparisons among samples and species within samples. Here we use a method (“DeSeq2”) initially developed for RNAseq data, to make quantifiable comparisons of expression possible among genes and gene pathways possible despite these biases. But this and other methods can only claim to make *relative* differences in community composition among samples (or expression of mRNA sequences) meaningful.

All of this to say, that as we understand Reviewer #1, they would like some comparison of the original diversity of fungi within the samples, and unfortunately this cannot be honestly recovered from the sample prep and bioinformatic pipeline. we would not trust any patterns in species richness from any stage of these pipelines, not due to any intellectual flaw in them but simply because these data and methods cannot produce that kind of data without large danger of false positives.

For more discussion of some of the above issues, and a list of useful sources, we have a preprinted discussion at < <https://doi.org/10.1101/184960> >.

**The figures are interesting, but in my opinion there are too many and they can be simplified.  I'd suggest just removing the outliers from the NMDS plots and stating as much in the caption (no need to present the outliers), or just present a discontinuous scale.**

I’ve reduced the total number of figures and simplified some of the remaining, multi-paneled figures. Some are moved to supplementary information.   
  
**I'm guessing that the average Fungal Ecology reader won't be familiar with the presentation of the PCNM vectors.  Please explain this in the figure legend (what do the size of the bubbles represent, what are the values?).**

Language around interpretation of PCNM vectors has been added.

**Reviewer #2: Thank you for the opportunity to review the paper by Thomas et al. It's generally a nicely done study that places endophytes in the context of spatial and environmental factors, hosts, and tissue types (wood, leaves). Generally the work is solid and the stats top-quality. Most of my suggestions are cosmetic.**

**- Please try to avoid 'nouns stacked up as modifiers' like this: leaf fungal endophyte amplicon library preparations**

We removed the examples of this we could find.

**- Can the description of the FDP be shortened given the existence of previously published information?**

Section has been shortened.

**- Where possible avoid single-sentence paragraphs (as in results)**

We removed the examples of this we could find.

**- The results regarding 'Environmental effects on endophyte community composition' are difficult for me to understand/read.**

We’ve tried to condense and clarify, We hoped this helped.

**- In general the authors do a nice job of not over-inferring causality from the observational study, but occasionally that still comes through (e.g., environmental effects on). I suggest softening the language here and throughout, perhaps to match that used in the text of the preceding paragraph (Host species is the strongest predictor'...).**

Agreed, changed.

**- In general the results are often described more in terms of statistics than biological inference. The authors are terrific biologists, so I would encourage them to read this from the biological perspective and, using their good judgment to not over-infer/overstate biological stories (they're great about this), try to help the average endophyte biologist get to the biology presented here.**

Agreed. This paper is heavy on statistics and light on biology. Unfortunately, we have to be honest about the limitations of the techniques. This paper is a report of results from a high-throughput ITS survey of fungal endophytes. With this kind of data, most ecological conclusions must be drawn from broad, community-dissimilarity-based patterns. With medium and large community data, conclusions often have to be gleaned using tools that are not familiar for many biologists – eigenanalyses of various sorts, and complex spatial tools. Other, simpler and more classical statistical techniques are typically very inappropriate. Another of the main drawbacks to this kind of survey is that inevitably many species remain anonymous. And often these kind of data are massively noisy and rarely produce a simple message. The result of all this is a type of study that can be very unsatisfying to most mycologists. Most of us would prefer to be counting mushrooms/stromata, culturing our organisms, etc. But there are some important and basic questions concerning fungi that are best addressed through these means. We should note that another report from this data is in-progress that focuses on patterns of Xylariaceae, on both the forest floor and in the canopy – the authors are pursuing more organisms-focused questions from this dataset! But here we are using this dataset to vet some broad theoretical ecological questions.

**- The summary comparison could be the first paragraph of the discussion instead of appended on the results.**

It is reviewed/summarized in the first section of the discussion. Hopefully the redundancy isn’t too much.

**- I'm not a huge fan of the trend to use 'catchy titles with colons' when a scientifically descriptive title might be more effective.**

Title is changed.

**- I suggest being a little careful in the discussion to not over-focus on the peculiarities of the FDP or this study in a way that really limits the scope of inference. The hill may be important indeed, but in the absence of a control or other context, does that broaden the study as written? Consider retooling here to help others appreciate and apply your work to improving their own. I also think that getting into neutral factors may be a bit far abroad for the study as presented here.**

Hilltop references have mostly been removed. We would prefer to keep in the mention of neutral processes. Discussion of neutral processes is inevitable when environmental data falls short of explaining spatial patterns, and we have added language to be straightforward about our inability to test such ideas. When we point out the importance of southwestern valley as the only region of the plot where a core may be establishing in leaves. this is an implicit reference to neutral processes. Regardless of the niches of our core microbes, we suggest that perhaps it is merely protection from change that is allowing the microbiome of a host to begin developing some local structure. It is useful to at least mention this, and to be open about the theoretical framework being referenced. We would love to devote more space to developing these concepts, but the discussion is probably already too lengthy.

**- I wonder if it would be useful for the nested diagram in Fig 1 to be proportional to otu or other measures.**

There may be a little confusion on figure numbers. Fig. 1 in both the revised and original manuscript is a physical map. Some clarification is needed here.

**- In the upper panel of Fig 5, any reason not to remove the outlier and re-analyze? Ditto with the odd host in Fig 6? (Recentering partially addresses this, but the analysis may be more informative if repeated w/ and w/o that species, with supplements used to good effect).**

As mentioned elsewhere, we had to rebuild the bioinforamtics pipeline, due to a loss of data. While redoing this we modified my contaminant removal process, using slightly more stringent method. This process resulted in cleaner NMS graphic in both cases, hopefully resolving this issue.

**- Interesting that the core mycobiome contains lots of common genera that show up all the time in endophyte studies. Maybe a comment on this?**

As mentioned, we had to rebuild the bioinformatic and statistical pipelines. This time, the high-thoughput taxonomic assignments become much less certain using UTAX, so we manually blasted the core OTUs against UNITE. This gave me finer control over the quality cutoffs for the match, and showed me that most of the assignments of the core OTUs were very low confidence, so we had to discard them. Many of the genera were removed from the taxonomy table. *Phyllostica* remained, however, so the reviewer’s comment is still pertinent. Specifics are in the scripts, and some language has been added to the conclusion.

**- Throughout, the language could be a little more linear. Nested clauses/etc. are sometimes a bit more prevalent than necessary.**

We’ve tried to improve this (but we can’t help it, at least one of the authors often thinks, - and speaks - in a very nested manner).

**- Where possible I think figures could be combined and/or set up in a way to more clearly emphasize the key, generalizable results of the paper. Supplements could be used effectively to focus the paper more.**

I’ve moved several figures to supplements, and removed others.

**- Overall the paper is nice, and I think the text would benefit from a 'step-back-and-read' to be sure the authors' key points are clear.**

Agreed, we have struggled with the process of interpreting the really large and noisy dataset into a single cohesive story without oversimplifying, but also without confusing the reader. So we erred on the side of confusing the reader. In these revisions we have simplified manuscript as best as we can, though it remains a statistically heavy paper. we hope the situation is somewhat improved by these edits.