THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

An introduction to phylosymbiosis

Shen Jean Lim and Seth R. Bordenstein

Article citation details
Proc. R. Soc. B **287**: 20192900.
http://dx.doi.org/10.1098/rspb.2019.2900

Review timeline

Original submission:	10 July 2019
1st revised submission:	11 December 2019
2nd revised submission:	21 January 2020
Final acceptance:	7 February 2020

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Review History

RSPB-2019-1539.R0 (Original submission)

Review form: Reviewer 1

Recommendation
Major revision is needed (please make suggestions in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Acceptable

General interest: Is the paper of sufficient general interest?
Excellent

Quality of the paper: Is the overall quality of the paper suitable?
Poor

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
Lim and Bordenstein have taken on an important challenge with this review, and one that I think is timely: providing a touchstone article precisely defining the concept of phylosymbiosis, reviewing the cases in which it has been documented, justifying the utility of the concept, and outlining how it can be used to guide research into the future. Use of the term is not without some controversy in the field. I find it to be a useful concept, but frequently find myself having to defend it to more skeptical colleagues. A cleary and carefully written article that methodically laid out precisely what constituted phylosymbiosis and why it is a useful term would make that job much easier, and I think would do a lot to improve research in host-associated microbiomes generally.

This is not that, at least not yet. I found the writing to be far too vague to be useful throughout much of the manuscript. For some of my phylosymbioskeptical colleagues, I worry that the manuscript as presented would only harden their conviction that the term is of little use. There is too much assertion *that* phylosymbiosis is useful, and far too little evidence for how and why. I think the paper would benefit tremendously from some very carefully considered and spelled-out thought experiments, perhaps even employing simulated or toy data, showing exactly how a specific measurement (the correlation between microbiome dissimilarity and host phylogeny) could be used to interpret or guide investigation of these complex systems.

One of the most problematic aspects of phylosymbiosis, for some, is that there are many possible underlying mechanisms to generate such a pattern, as well as many possible ways to measure it. This needs to be tackled more directly. Some of my favorite papers on the subject, such as Groussin et al, have used this to their advantage by making explicit connections between the different processes that might be underlying phylosymbiotic patterns and their predicted effects on different ways of measuring these patterns. I understand that in trying to write an "introduction," the authors might be wary of getting too far into the weeds of details. But absent that kind of specificity, I worry that the authors' advice to report measurements of phylosymbiosis from across a wide range of parameter values will sound too much like throwing as much as possible at a wall to see what sticks. I think in its best possible realization, this article would engage very carefully with the implications of these details to help convey to a reader why and how a concept of phylosymbiosis has utility.

Below are specific comments. I apologize for the very long list, which I know will seem onerous to address. Please be assured that they come from a sincere desire to improve the manuscript,

and a great desire to see it published. I will restate that I think such a paper could be of great value to the field. Good luck!

51-54: This doesn't seem clear or specific enough of a definition to me -- wouldn't 'associations between two or more organisms' without temporal or categorical restriction encompass *any* association between organisms? Is that restrictive enough to be considered a useful and clear definition?

80: The use of the phrasing “a host phylogenetic effect on" seems problematic to me, as it implies a causal relationship from the phylogeny to the microbial community. I understand that in some contexts "a phylogenetic effect" can be used as statistical shorthand when talking about correlations, but given that this is meant to be an explanatory piece, I worry about implying causation so early on.

95-127: This section feels a little disjointed—the examples read a little like a list, with little to provide context or a narrative for the reader. These three paragraphs are followed by a final paragraph exploring potential adaptive significance of host-specific microbiomes. Is this part of the “history of phylosymiosis?” Why?

101: I would drop "randomized and non-randomized" here and simply say "comparisons of beta diversity-derived clusters with mammalian phylogenies...". The randomization is how the strength of the comparison was assessed, but without additional context it might seem confusing.

116: Is "functional genetic" the best term here? Some readers, particularly from more ecological or natural history backgrounds, may not have an immediate sense of what is implied by “functional genetics" in this context. Why not simply say "The first study linking phylosymbiotic patterns to the function of specific host genes”?

132: “Microbiome studies have demonstrated that resident microbes can preferentially colonize some host taxa, likely through host-specific... colonization (20)." This seems like a truism as written?

134: “through non-immune and immune factors" would seem to just about cover the possible types of factors... is it necessary to specify?

135: What is "functional phylosymbiosis”?

136-137: Why specifically might hybridization between closely related host species disrupt host-microbiome associations?

143-144: Is this really a fair generalization of the results of the Nasonia experiments cited? I would agree that the experiment demonstrates how microbes can be an integral component of hybrid sterility and post-zygotic isolation, and that this in turn can help explain how phylosymbiotic patterns might arise. But is it really the phylosymbiotic-ness of the community that's leading to adverse consequences?

165: This is just a linguistic quibble, but can host ancestors split into different but _un_related genera?

175-178: Earlier, you state (and I think accurately) that “a positive association between host phylogenetic and microbial community relationships does not a priori imply a shared and ancestral evolutionary history... it may or may not reflect long-term associations or co-adaptations.” Here, though, you state that "specialized host-microbe associations indicate that hosts are adapted to their native microbiomes.” These statements seem in some conflict!

Also, I don't see how these examples indicate that the given phylosymbiotic patterns arose via

natural selection and not neutral evolution. Neutral drift can absolutely lead to phenotypic consequences even in the absence of selection.

193: What are "phylosymbiosis outcomes”?

195-196: This statement seems like it could be expanded. How is 'stochastically' being used here? One could imagine stochastic processes–such as geographically isolated host populations exposed to different microbial metacommunities–that would result in a pattern of phylosymbiosis.

To me, it would be more useful here to explicitly consider a couple different scenarios that might produce a pattern of phylosymbiosis, rather than simply asserting a statement like this.

204-205: Where does the recommendation for an N of 4 come from? 24 is not a tremendous number of possible permutations, meaning that any analysis with just 4 taxa is going to be extremely underpowered.

218: This section makes no mention of the quandary of how to represent multiple individuals from the same species in an assessment of phylosymbiosis, which was noted by Mazel et al to have the potential to vastly inflate false positives.

220: Similar to above, where does a recommendation of ten samples come from? What is the statistical estimate whose confidence is being maximized here? Increasing the number of samples per species will improve the estimate of that taxon's microbiome, but has no bearing on the power of the estimate of phylosymbiosis itself, which is ultimately constrained by the number of taxa assessed.

238-239: Why should quantitative metrics be preferred? For that matter, why should *any* type of metric be preferred? They are all measurements of community dissimilarity that reflect different aspects of that dissimilarity. Quantitative metrics, while less sensitive to rare taxa, can be overwhelmed by variation in a few very abundant ones. Which is appropriate for a particular study?

247-249: How should researchers interpret the inevitable differences among these parameter combinations? This also raises the specter of multiple testing and specificity, which should also be discussed in the context of recommended size of the taxon sample. How should we interpret when one of twelve tested parameter combinations reveals phylosymbiosis?

255: How does distinguishability relate to phylosymbiosis? Is distinguishability a prerequisite for phylosymbiosis?

280: Microbiome similarity dendrograms are also somewhat arbitrary (why use UPGMA clustering, instead of one of the many alternatives?), and especially when the number of samples analyzed increases, can have very unstable topologies. How is this to be interpreted in the context of phylosymbiosis?

296-298: I don't quite understand this sentence. Is it meant to be specific to matrix correlation methods? Also to topology-based methods?

Additionally, Mazel et al found matrix correlation methods to have significant advantages over topology-based approaches using their simulated data benchmarks. This seems relevant to mention here, no?

342: "gender" should be replaced with "sex" in this context.

344-347: I'm having trouble with this sentence. What is a cluster of metagenomes? And is

"microbial taxonomic order" meant to be “host taxonomic order”? (It would be a truism that OTUs would vary with microbial taxonomy, no?) Why would covariance of bacterial relative abundance and host genes imply a functional relationship? It seems equally possible that this sort of correlation could be indirect.

377-379: Is correlation of microbial alpha diversity with host phylogeny considered a facet of phylosymbiosis? If so, it should probably be explored more explicitly early on.

379: I worry that referencing per-lineage cophylogeny here will tend to muddle the distinction of 'what phylosymbiosis is' to a new reader. Having examples of cophylogeny or correlations with alpha diversity mixed in with these examples might be counterproductive -- would it be worth instead compartmentalizing them in a brief additional section on other phylogenetic correlations?

383: ‘bacterial bacteriomes' is redundant

397-398: is it reasonable to discuss 'non-statistically significant phylosymbiotic correlations’?

402-428: This section should be one of the highlights of this review—anyone skeptical of the utility of the term and concept is likely to skip to this heading to find out what IS the significance of phylosymbiosis. I'm worried that their conclusion might be “there is none.” Most of what is presented are assertions *that* phylosymbiosis is or provides something, rather than convincing examples or evidence for *why* and *how*.

414-415: This section would benefit tremendously from some concrete examples. I'm having a hard time understanding how a pattern can provide eco-evolutionary predictions. This is especially true for something like phylosymbiosis, which is a pattern that can conceivably be generated by quite a large number of rather unrelated processes.

417-419: As with the previous sentence, this one is quite difficult for me to meaningfully parse. (What is an ecological mechanism in this context? What would make such a mechanism intrinsic, and intrinsic to what?)

421-428: How, exactly, does phylosymbiosis contribute to this growing school of thought? Please be specific. What precisely can a measure of a pattern like phylosymbiosis do to help us disentangle these diverse interactions—to 'determine the magnitude of each of these traits’?

451-452: “species conservation" seems to come a little out of left field here. Is this referring to the conservation of symbiont species across evolutionary time within lineages, or the conservation of threatened species?

458: "rules and themes”: what would be an example of a “rule" pertaining to phylosymbiosis, and how would that differ from a "theme”? Given that phylosymbiosis is just a pattern, why exactly would a standardized workflow for its measurement lead directly to establishment of such "rules”?

461-462: I think this concluding sentence will be useful in reassessing the rest of the paper. I completely agree that we need a much more thorough understanding of the mechanisms underlying host-microbe associations. I *also* agree that documentation of high-level patterns, like phylosymbiosis, can play a really useful role in the investigation of such processes. But much more work needs to be done in the piece to very carefully draw explicit connections from the observation of pattern to how, precisely, such an observation is likely to guide investigation of process.

Review form: Reviewer 2

Recommendation
Reject – article is not of sufficient interest (we will consider a transfer to another journal)

Scientific importance: Is the manuscript an original and important contribution to its field?
Poor

General interest: Is the paper of sufficient general interest?
Acceptable

Quality of the paper: Is the overall quality of the paper suitable?
Marginal

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
Yes

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
Yes

Is it clear?
Yes

Is it adequate?
No

Do you have any ethical concerns with this paper?
No

Comments to the Author
The authors present a review of host-microbiome "co-phylogeny" studies (i.e., studies comparing the phylogenies of hosts and with the phylogenies of their microbiota) within a framework informed by the concept of "phylosymbiosis". The phylosymbiosis framework is a very useful one for host-symbiont studies, Unfortunately, the paper (and probably the concept) needs some clarification and revisions before the framework is ready to be shared with the community. The authors promise presentation of a framework for studies of phylosymbiosis, with a review that "serves as a gateway to experimental, conceptual, and quantitative themes". As it stands, this paper falls short of this goal. In particular it fails to serve as a standalone description of the phylosybiosis framework: one has to have read, in depth, the Brooks et al. 2017 paper referenced many times in this work, to actually even know the statistics as well as predictions of "phylosybiosis", and having done that, it is not clear to me how much additional insight this paper contributes to the discussion.

Most of the problems lie in the first half the paper, where the phylosymbiosis framework is

(purportedly) presented along with various methods for assessing phylosybiosis. This presentation is woefully inadequate and very disappointing, lacking in conceptual and statistical detail. Most of the methods are presented simply by name, with no value-added interpretation, explanation, or insight into their mechanics. Beyond the relatively narrow scope of statistical methodologies, however, in general the presentation of the phylosymbiosis framework is extremely disappointing. It is insufficient in detail to really understand the framework, and there is no additional insight or interpretation provided to make it worthwhile reading this paper after having read some of the primary works it itself references (such as the Brooks et al. 2017 paper).

- [LL 78-93] Here the authors qualitatively define "phylosymbiosis". The authors define phylosymbiosis as "microbial community relationships that recapitulate the history of the host". The authors then go on to say that phylosymbiosis is a "pattern observed in a snapshot in time and space, and may or may not reflect long-term associations or co-adaptations that can be subsequently evaluated by empirical analyses". Many issues with this paper could have been avoided if the authors went on to recapitulate the major points, predictions, and quantitative concepts from the Brooks et al., 2017 paper here as well, to clearly and explicitly define the phylosymbiosis framework.

- [LL. 81-83] The authors characterize "phylogeography" as "the study of evolutionary processes that shape geographic or ecological distributions of organisms".
  Phylogeography is understood today as the study of the association between geographical and phylogenetic relationships, typically through the influence of the former on driving historical demographic processes (or events) that in turn structure the latter.
  In this sense, the analogy with "phylosymbiosis" the authors is actually reverse in causality.
  In phylosymbiosis, the "phylo" part (i.e., host's phylogenetic structure) is the casual component, informing the symbiote relationships, whereas in phylogeography, the "phylo" part is the dependent component, being informed by the geography.
  This is a minor, and, in the larger picture, entirely inconsequential point, I know, but I mention it here nonetheless.

- [LL. 85-88] What is the difference between "host phylogenetic effect" (or definition of this) vs. "reciprocal evolutionary genetic changes" or "ancestral splitting of host and symbiotic lineages"? Without a clear definition of the term to the contrary, I would understand "host phylogenetic effect" as possibly including the latter two.

- [LL. 101-103] "suggested a weak phylosymbiotic association"; as with "host phylogenetic effect" above, this term really should be defined before assertions or statements about it are made. In this case, the qualifier "weak" implies a quantitative aspect that needs to be explicitly described beforehand.

- [LL. 135-135] "functional phylosymbiosis" -- the term is used here with no explicit definition. The reader might be able to infer its meaning from context, but (a) I'm not sure if two readers would agree; (b) in this case of this particular reader at least, the meaning I inferred is vague enough that I would not be able to explain it clearly to a third party; (c) in a work introducing this framework, this term should be explicitly defined and not left as an exercise to the reader.

- [LL 135-148] I am afraid that I could not follow this argument.
  - I was led to expect that I would be shown two ways to empirically test for "functional phylosymbiosis" (or what I inferred "functional symbiosis" to mean) by the statement "functional symbiosis can be evaluated empirically in at least two ways".
  - To follow through on this, the authors need to first clearly describe what the predictions of functional phylosymbiosis are, and then describe how empirical data can be tested for patterns conforming to these predictions. In addition, to be convincing, the authors need to demonstrate that there are no alternate explanations for these patterns or otherwise taken into account alternate explanations in their tests.
  - However, instead, for the first of their two evalations the authors present one possible

outcome of hybridization (host-microbiome disruption in parental species) and go on to possible outcomes of this, in turn, in F2 generation in one particular system (Nasonia). The authors then conclude that "non-phylosymbiotic relationships can lead to adverse functional and evolutionary consequences over time".

- First of all, the authors assertion that host-microbiom disruption in parental species as an outcome of hybridization needs to be demonstrated, at least conceptually or otherwise with reference to previous work.
- More importantly, the authors have merely provided an interpretation of a pattern with reference to a "non-phylosymbiotic relationships" without actually showing how this is a prediction of functional phylosymbiosis let alone an exclusive prediction or serve as a satisfactory test for it.
- The second part of this argument states that "microbiome transplant experiments between related species/lineages will lead to host fitness reductions ....". This may be the case, but it is unclear to me how this provides an assessment of "phylosymbiosis", at least as defined by the authors, i.e., "microbial community relationships that recapitulate the history of the host".
- Furthermore, even if this pattern were to be consistent or (not inconsistent) with functional symbiosis, it seems that this still falls short of a diagnosis of functional symbiosis (which is what I took the authors to mean when they state they are going to present an evaluation)

- [LL 151-153] This was a very disappointing section, where the authors again promise something without adequately delivering it. Authors begin by stating that they will propose an "initial, methodological workflow to statistically evaluate its [phylosymbiosis] strength and significance". Unfortunately, description of this workflow or its associated statistics are not seen. The authors go on to describe a particular experiment and report the secondary conclusions/interpretations from other papers *without* any actual formal description of the tests, statistics, criteria, or any other details of the workflow. I understand that this might have been described in the previous work, but (a) the authors promise a description in *this* paper; (b) as a review paper that claims to formally introduce the framework, such a description does indeed belong in this paper; (c) without a formal description of the workflow, if they are just going to refer or rehash the results of another paper, without any added insight, then I do not see the point of this paper.

- [LL 234-235] In a review paper establishing a framework like this, the authors should at the very least present references to these measures.

- [LL 238-239] "are preferred estimators" --- clarify reasoning

- [LL 276-280] The authors refer to a set of scripts to evaluate the statistical significance of the RF and Matching Cluster distances, and, again, as throughout this paper, defer any explanation, description, and justification of the actual statistics to the frequently-cited Brooks et al. (2017) paper. Again, a reader of this review would not understand much without reading the Brooks et al. 2017 paper, and, conversely, a reader who has read that paper would not gain much at all from reading this paper.

- [LL 276-280] Furthermore, the scripts referenced to do this calculations are poorly documented. The program design is extremely primitive (with, for e.g., paths hard-coded into the scripts and users expected to modify the scripts from run to run for different projects), and there is no indication of any test suites to check for validity of input/output. Furthermore, some of the scripts are simply broken (e.g., one the scripts has the boilerplate copyright text uncommented, resulting in the Python interpreter attempting to execute them). I am aware that the authors themselves are not the primary maintainer of the scripts, nor are the scripts being presented in this paper. However, the authors do reference the scripts as the primary tool by which they evaluate the degree of phylosymbiosis and recommend these scripts to the user.

- [LL 288-289] "comparing results with those produced by matrix correlation methods". The authors should provide some discussion on the reasons that these methods differ in their

approaches and assumptions, and, in particular, what might be captured by one method but not the other. We should have a sophisticated enough appreciation of our statistical approaches not just to throw a million analysis at a problem and hope that they all converge. We pick our methods carefully to address the questions and problems at hand, and when using multiple methods with pick them so that they complement each other's strengths. What do matrix correlation methods reveal that might be missed by topological congruence methods? What

- [LL 291-298] This description of matrix correlation methods is disappointingly inadequate for the declared scope of this review paper. A laundry list of methods are mentioned by name as having "been implemented in phylosymbiosis studes" with, again, absolutely no description, characterization, interpretation, or any insight about these methods at all.

Review form: Reviewer 3

Recommendation
Accept with minor revision (please list in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Excellent

General interest: Is the paper of sufficient general interest?
Excellent

Quality of the paper: Is the overall quality of the paper suitable?
Excellent

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
This review is very well written, it summarizes the main evidence supporting the phylosymbioisis concept across diverse species and with specific examples on how phylosymbioisis shapes host-microbiome associations and fitness across living systems. It also summarizes methodological steps to characterize phylosymbiosis.

I only have 2 suggestions on additional minor additions (sections) that would make a more complete review.

1. How are current meta-OMIC techniques to characterize the microbiome and that denote functional rather than taxonomic modules (metagenomics- metabolomics) fit in the phylosymbioisis concept? Most analyses of these type have been based on taxonomic microbiome surveys. Also, would we expect microbiome functionality (genes/pathways/metabolites) to be less constrained by host phylogeny than microbiome taxa/phylogeny?

2. Along these lines, a section with specific examples in which phylosymbiosis has not been supported or in which host phylogenetic patterns does not support congruent microbiomes could be important. Specifically, the importance of shared dietary niche in different (less related) hosts, in shaping or rejecting phylosymbiosis could be discussed.

Decision letter (RSPB-2019-1539.R0)

05-Sep-2019

Dear Dr Bordenstein:

I am writing to inform you that your manuscript RSPB-2019-1539 entitled "An Introduction to Phylosymbiosis" has, in its current form, been rejected for publication in Proceedings B.

This action has been taken on the advice of referees, who have recommended that very substantial revisions are necessary. With this in mind we would be happy to consider a resubmission, provided the comments of the referees are fully addressed. However please note that this is not a provisional acceptance.

The resubmission will be treated as a new manuscript. However, we will approach the same reviewers if they are available and it is deemed appropriate to do so by the Editor. Please note that resubmissions must be submitted within six months of the date of this email. In exceptional circumstances, extensions may be possible if agreed with the Editorial Office. Manuscripts submitted after this date will be automatically rejected.

Please find below the comments made by the referees, not including confidential reports to the Editor, which I hope you will find useful. If you do choose to resubmit your manuscript, please upload the following:

1) A ‘response to referees’ document including details of how you have responded to the comments, and the adjustments you have made.
2) A clean copy of the manuscript and one with 'tracked changes' indicating your 'response to referees' comments document.
3) Line numbers in your main document.

To upload a resubmitted manuscript, log into http://mc.manuscriptcentral.com/prsb and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with

Decisions." Under "Actions," click on "Create a Resubmission." Please be sure to indicate in your cover letter that it is a resubmission, and supply the previous reference number.

Sincerely,
Professor Hans Heesterbeek, Editor
mailto: proceedingsb@royalsociety.org

Associate Editor's Comments to Author:

Thank you for submitting your work for the special issue. The manuscript has now been read by myself and three reviewers and we all see merit in the work, which is a timely piece covering a still controversial topic. As you will see, the reviewers have offered a substantial number of suggestions for improvement and generally felt that the need for a piece like this in the field is great, but that the manuscript has some way to go before it can act as the strong contribution it has the potential to be. In particular, the reviewers felt that more specific statements about how the concept can be usefully applied was warranted, and that overall more clearly spelled out definitions and ideas would help readers navigate this complex topic. I agree with the general sentiment and look forward to receiving a revised manuscript that takes these thoughtful suggestions into account. Thank you again.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)

Lim and Bordenstein have taken on an important challenge with this review, and one that I think is timely: providing a touchstone article precisely defining the concept of phylosymbiosis, reviewing the cases in which it has been documented, justifying the utility of the concept, and outlining how it can be used to guide research into the future. Use of the term is not without some controversy in the field. I find it to be a useful concept, but frequently find myself having to defend it to more skeptical colleagues. A cleary and carefully written article that methodically laid out precisely what constituted phylosymbiosis and why it is a useful term would make that job much easier, and I think would do a lot to improve research in host-associated microbiomes generally.

This is not that, at least not yet. I found the writing to be far too vague to be useful throughout much of the manuscript. For some of my phylosymbioskeptical colleagues, I worry that the manuscript as presented would only harden their conviction that the term is of little use. There is too much assertion *that* phylosymbiosis is useful, and far too little evidence for how and why. I think the paper would benefit tremendously from some very carefully considered and spelled-out thought experiments, perhaps even employing simulated or toy data, showing exactly how a specific measurement (the correlation between microbiome dissimilarity and host phylogeny) could be used to interpret or guide investigation of these complex systems.

One of the most problematic aspects of phylosymbiosis, for some, is that there are many possible underlying mechanisms to generate such a pattern, as well as many possible ways to measure it. This needs to be tackled more directly. Some of my favorite papers on the subject, such as Groussin et al, have used this to their advantage by making explicit connections between the different processes that might be underlying phylosymbiotic patterns and their predicted effects on different ways of measuring these patterns. I understand that in trying to write an "introduction," the authors might be wary of getting too far into the weeds of details. But absent that kind of specificity, I worry that the authors' advice to report measurements of phylosymbiosis from across a wide range of parameter values will sound too much like throwing

as much as possible at a wall to see what sticks. I think in its best possible realization, this article would engage very carefully with the implications of these details to help convey to a reader why and how a concept of phylosymbiosis has utility.

Below are specific comments. I apologize for the very long list, which I know will seem onerous to address. Please be assured that they come from a sincere desire to improve the manuscript, and a great desire to see it published. I will restate that I think such a paper could be of great value to the field. Good luck!

51-54: This doesn't seem clear or specific enough of a definition to me -- wouldn't 'associations between two or more organisms' without temporal or categorical restriction encompass *any* association between organisms? Is that restrictive enough to be considered a useful and clear definition?

80: The use of the phrasing “a host phylogenetic effect on" seems problematic to me, as it implies a causal relationship from the phylogeny to the microbial community. I understand that in some contexts "a phylogenetic effect" can be used as statistical shorthand when talking about correlations, but given that this is meant to be an explanatory piece, I worry about implying causation so early on.

95-127: This section feels a little disjointed—the examples read a little like a list, with little to provide context or a narrative for the reader. These three paragraphs are followed by a final paragraph exploring potential adaptive significance of host-specific microbiomes. Is this part of the “history of phylosymiosis?” Why?

101: I would drop "randomized and non-randomized" here and simply say "comparisons of beta diversity-derived clusters with mammalian phylogenies...". The randomization is how the strength of the comparison was assessed, but without additional context it might seem confusing.

116: Is "functional genetic" the best term here? Some readers, particularly from more ecological or natural history backgrounds, may not have an immediate sense of what is implied by “functional genetics" in this context. Why not simply say "The first study linking phylosymbiotic patterns to the function of specific host genes”?

132: “Microbiome studies have demonstrated that resident microbes can preferentially colonize some host taxa, likely through host-specific... colonization (20)." This seems like a truism as written?

134: “through non-immune and immune factors" would seem to just about cover the possible types of factors... is it necessary to specify?

135: What is "functional phylosymbiosis”?

136-137: Why specifically might hybridization between closely related host species disrupt host-microbiome associations?

143-144: Is this really a fair generalization of the results of the Nasonia experiments cited? I would agree that the experiment demonstrates how microbes can be an integral component of hybrid sterility and post-zygotic isolation, and that this in turn can help explain how phylosymbiotic patterns might arise. But is it really the phylosymbiotic-ness of the community that's leading to adverse consequences?

165: This is just a linguistic quibble, but can host ancestors split into different but _un_related genera?

175-178: Earlier, you state (and I think accurately) that “a positive association between host

phylogenetic and microbial community relationships does not a priori imply a shared and ancestral evolutionary history... it may or may not reflect long-term associations or co-adaptations." Here, though, you state that "specialized host-microbe associations indicate that hosts are adapted to their native microbiomes." These statements seem in some conflict!

Also, I don't see how these examples indicate that the given phylosymbiotic patterns arose via natural selection and not neutral evolution. Neutral drift can absolutely lead to phenotypic consequences even in the absence of selection.

193: What are "phylosymbiosis outcomes"?

195-196: This statement seems like it could be expanded. How is 'stochastically' being used here? One could imagine stochastic processes–such as geographically isolated host populations exposed to different microbial metacommunities–that would result in a pattern of phylosymbiosis.

To me, it would be more useful here to explicitly consider a couple different scenarios that might produce a pattern of phylosymbiosis, rather than simply asserting a statement like this.

204-205: Where does the recommendation for an N of 4 come from? 24 is not a tremendous number of possible permutations, meaning that any analysis with just 4 taxa is going to be extremely underpowered.

218: This section makes no mention of the quandary of how to represent multiple individuals from the same species in an assessment of phylosymbiosis, which was noted by Mazel et al to have the potential to vastly inflate false positives.

220: Similar to above, where does a recommendation of ten samples come from? What is the statistical estimate whose confidence is being maximized here? Increasing the number of samples per species will improve the estimate of that taxon's microbiome, but has no bearing on the power of the estimate of phylosymbiosis itself, which is ultimately constrained by the number of taxa assessed.

238-239: Why should quantitative metrics be preferred? For that matter, why should *any* type of metric be preferred? They are all measurements of community dissimilarity that reflect different aspects of that dissimilarity. Quantitative metrics, while less sensitive to rare taxa, can be overwhelmed by variation in a few very abundant ones. Which is appropriate for a particular study?

247-249: How should researchers interpret the inevitable differences among these parameter combinations? This also raises the specter of multiple testing and specificity, which should also be discussed in the context of recommended size of the taxon sample. How should we interpret when one of twelve tested parameter combinations reveals phylosymbiosis?

255: How does distinguishability relate to phylosymbiosis? Is distinguishability a prerequisite for phylosymbiosis?

280: Microbiome similarity dendrograms are also somewhat arbitrary (why use UPGMA clustering, instead of one of the many alternatives?), and especially when the number of samples analyzed increases, can have very unstable topologies. How is this to be interpreted in the context of phylosymbiosis?

296-298: I don't quite understand this sentence. Is it meant to be specific to matrix correlation methods? Also to topology-based methods?

Additionally, Mazel et al found matrix correlation methods to have significant advantages over

topology-based approaches using their simulated data benchmarks. This seems relevant to mention here, no?

342: "gender" should be replaced with "sex" in this context.

344-347: I'm having trouble with this sentence. What is a cluster of metagenomes? And is "microbial taxonomic order" meant to be "host taxonomic order"? (It would be a truism that OTUs would vary with microbial taxonomy, no?) Why would covariance of bacterial relative abundance and host genes imply a functional relationship? It seems equally possible that this sort of correlation could be indirect.

377-379: Is correlation of microbial alpha diversity with host phylogeny considered a facet of phylosymbiosis? If so, it should probably be explored more explicitly early on.

379: I worry that referencing per-lineage cophylogeny here will tend to muddle the distinction of 'what phylosymbiosis is' to a new reader. Having examples of cophylogeny or correlations with alpha diversity mixed in with these examples might be counterproductive -- would it be worth instead compartmentalizing them in a brief additional section on other phylogenetic correlations?

383: 'bacterial bacteriomes' is redundant

397-398: is it reasonable to discuss 'non-statistically significant phylosymbiotic correlations'?

402-428: This section should be one of the highlights of this review—anyone skeptical of the utility of the term and concept is likely to skip to this heading to find out what IS the significance of phylosymbiosis. I'm worried that their conclusion might be "there is none." Most of what is presented are assertions *that* phylosymbiosis is or provides something, rather than convincing examples or evidence for *why* and *how*.

414-415: This section would benefit tremendously from some concrete examples. I'm having a hard time understanding how a pattern can provide eco-evolutionary predictions. This is especially true for something like phylosymbiosis, which is a pattern that can conceivably be generated by quite a large number of rather unrelated processes.

417-419: As with the previous sentence, this one is quite difficult for me to meaningfully parse. (What is an ecological mechanism in this context? What would make such a mechanism intrinsic, and intrinsic to what?)

421-428: How, exactly, does phylosymbiosis contribute to this growing school of thought? Please be specific. What precisely can a measure of a pattern like phylosymbiosis do to help us disentangle these diverse interactions—to 'determine the magnitude of each of these traits'?

451-452: "species conservation" seems to come a little out of left field here. Is this referring to the conservation of symbiont species across evolutionary time within lineages, or the conservation of threatened species?

458: "rules and themes": what would be an example of a "rule" pertaining to phylosymbiosis, and how would that differ from a "theme"? Given that phylosymbiosis is just a pattern, why exactly would a standardized workflow for its measurement lead directly to establishment of such "rules"?

461-462: I think this concluding sentence will be useful in reassessing the rest of the paper. I completely agree that we need a much more thorough understanding of the mechanisms underlying host-microbe associations. I *also* agree that documentation of high-level patterns, like phylosymbiosis, can play a really useful role in the investigation of such processes. But much more work needs to be done in the piece to very carefully draw explicit connections from the

observation of pattern to how, precisely, such an observation is likely to guide investigation of process.

Referee: 2

Comments to the Author(s)
The authors present a review of host-microbiome "co-phylogeny" studies (i.e., studies comparing the phylogenies of hosts and with the phylogenies of their microbiota) within a framework informed by the concept of "phylosymbiosis". The phylosymbiosis framework is a very useful one for host-symbiont studies, Unfortunately, the paper (and probably the concept) needs some clarification and revisions before the framework is ready to be shared with the community. The authors promise presentation of a framework for studies of phylosymbiosis, with a review that "serves as a gateway to experimental, conceptual, and quantitative themes". As it stands, this paper falls short of this goal. In particular it fails to serve as a standalone description of the phylosybiosis framework: one has to have read, in depth, the Brooks et al. 2017 paper referenced many times in this work, to actually even know the statistics as well as predictions of "phylosybiosis", and having done that, it is not clear to me how much additional insight this paper contributes to the discussion.

Most of the problems lie in the first half the paper, where the phylosymbiosis framework is (purportedly) presented along with various methods for assessing phylosybiosis. This presentation is woefully inadequate and very disappointing, lacking in conceptual and statistical detail. Most of the methods are presented simply by name, with no value-added interpretation, explanation, or insight into their mechanics. Beyond the relatively narrow scope of statistical methodologies, however, in general the presentation of the phylosymbiosis framework is extremely disappointing. It is insufficient in detail to really understand the framework, and there is no additional insight or interpretation provided to make it worthwhile reading this paper after having read some of the primary works it itself references (such as the Brooks et al. 2017 paper).

- [LL 78-93] Here the authors qualitatively define "phylosymbiosis". The authors define phylosymbiosis as "microbial community relationships that recapitulate the history of the host". The authors then go on to say that phylosymbiosis is a "pattern observed in a snapshot in time and space, and may or may not reflect long-term associations or co-adaptations that can be subsequently evaluated by empirical analyses". Many issues with this paper could have been avoided if the authors went on to recapitulate the major points, predictions, and quantitative concepts from the Brooks et al., 2017 paper here as well, to clearly and explicitly define the phylosymbiosis framework.

- [LL. 81-83] The authors characterize "phylogeography" as "the study of evolutionary processes that shape geographic or ecological distributions of organisms".
  Phylogeography is understood today as the study of the association between geographical and phylogenetic relationships, typically through the influence of the former on driving historical demographic processes (or events) that in turn structure the latter.
  In this sense, the analogy with "phylosymbiosis" the authors is actually reverse in causality.
  In phylosymbiosis, the "phylo" part (i.e., host's phylogenetic structure) is the casual component, informing the symbiote relationships, whereas in phylogeography, the "phylo" part is the dependent component, being informed by the geography.
  This is a minor, and, in the larger picture, entirely inconsequential point, I know, but I mention it here nonetheless.

- [LL. 85-88] What is the difference between "host phylogenetic effect" (or definition of this) vs. "reciprocal evolutionary genetic changes" or "ancestral splitting of host and symbiotic lineages"? Without a clear definition of the term to the contrary, I would understand "host phylogenetic effect" as possibly including the latter two.

- [LL. 101-103] "suggested a weak phylosymbiotic association"; as with "host phylogenetic effect" above, this term really should be defined before assertions or statements about it are made. In this case, the qualifier "weak" implies a quantitative aspect that needs to be explicitly described beforehand.

- [LL. 135-135] "functional phylosymbiosis" -- the term is used here with no explicit definition. The reader might be able to infer its meaning from context, but (a) I'm not sure if two readers would agree; (b) in this case of this particular reader at least, the meaning I inferred is vague enough that I would not be able to explain it clearly to a third party; (c) in a work introducing this framework, this term should be explicitly defined and not left as an exercise to the reader.

- [LL 135-148] I am afraid that I could not follow this argument.
  - I was led to expect that I would be shown two ways to empirically test for "functional phylosymbiosis" (or what I inferred "functional symbiosis" to mean) by the statement "functional symbiosis can be evaluated empirically in at least two ways".
  - To follow through on this, the authors need to first clearly describe what the predictions of functional phylosymbiosis are, and then describe how empirical data can be tested for patterns conforming to these predictions. In addition, to be convincing, the authors need to demonstrate that there are no alternate explanations for these patterns or otherwise taken into account alternate explanations in their tests.
  - However, instead, for the first of their two evalations the authors present one possible outcome of hybridization (host-microbiome disruption in parental species) and go on to possible outcomes of this, in turn, in F2 generation in one particular system (Nasonia). The authors then conclude that "non-phylosymbiotic relationships can lead to adverse functional and evolutionary consequences over time".
    - First of all, the authors assertion that host-microbiom disruption in parental species as an outcome of hybridization needs to be demonstrated, at least conceptually or otherwise with reference to previous work.
    - More importantly, the authors have merely provided an interpretation of a pattern with reference to a "non-phylosymbiotic relationships" without actually showing how this is a prediction of functional phylosymbiosis let alone an exclusive prediction or serve as a satisfactory test for it.
  - The second part of this argument states that "microbiome transplant experiments between related species/lineages will lead to host fitness reductions ....". This may be the case, but it is unclear to me how this provides an assessment of "phylosymbiosis", at least as defined by the authors, i.e., "microbial community relationships that recapitulate the history of the host".
  - Furthermore, even if this pattern were to be consistent or (not inconsistent) with functional symbiosis, it seems that this still falls short of a diagnosis of functional symbiosis (which is what I took the authors to mean when they state they are going to present an evaluation)

- [LL 151-153] This was a very disappointing section, where the authors again promise something without adequately delivering it. Authors begin by stating that they will propose an "initial, methodological workflow to statistically evaluate its [phylosymbiosis] strength and significance". Unfortunately, description of this workflow or its associated statistics are not seen. The authors go on to describe a particular experiment and report the secondary conclusions/interpretations from other papers *without* any actual formal description of the tests, statistics, criteria, or any other details of the workflow. I understand that this might have been described in the previous work, but (a) the authors promise a description in *this* paper; (b) as a review paper that claims to formally introduce the framework, such a description does indeed belong in this paper; (c) without a formal description of the workflow, if they are just going to refer or rehash the results of another paper, without any added insight, then I do not see the point of this paper.

- [LL 234-235] In a review paper establishing a framework like this, the authors should at the very least present references to these measures.

- [LL 238-239] "are preferred estimators" --- clarify reasoning

- [LL 276-280] The authors refer to a set of scripts to evaluate the statistical significance of the RF and Matching Cluster distances, and, again, as throughout this paper, defer any explanation, description, and justification of the actual statistics to the frequently-cited Brooks et al. (2017) paper. Again, a reader of this review would not understand much without reading the Brooks et al. 2017 paper, and, conversely, a reader who has read that paper would not gain much at all from reading this paper.

- [LL 276-280] Furthermore, the scripts referenced to do this calculations are poorly documented. The program design is extremely primitive (with, for e.g., paths hard-coded into the scripts and users expected to modify the scripts from run to run for different projects), and there is no indication of any test suites to check for validity of input/output. Furthermore, some of the scripts are simply broken (e.g., one the scripts has the boilerplate copyright text uncommented, resulting in the Python interpreter attempting to execute them). I am aware that the authors themselves are not the primary maintainer of the scripts, nor are the scripts being presented in this paper. However, the authors do reference the scripts as the primary tool by which they evaluate the degree of phylosymbiosis and recommend these scripts to the user.

- [LL 288-289] "comparing results with those produced by matrix correlation methods". The authors should provide some discussion on the reasons that these methods differ in their approaches and assumptions, and, in particular, what might be captured by one method but not the other. We should have a sophisticated enough appreciation of our statistical approaches not just to throw a million analysis at a problem and hope that they all converge. We pick our methods carefully to address the questions and problems at hand, and when using multiple methods with pick them so that they complement each other's strengths. What do matrix correlation methods reveal that might be missed by topological congruence methods? What

- [LL 291-298] This description of matrix correlation methods is disappointingly inadequate for the declared scope of this review paper. A laundry list of methods are mentioned by name as having "been implemented in phylosymbiosis studes" with, again, absolutely no description, characterization, interpretation, or any insight about these methods at all.

Referee: 3

Comments to the Author(s)
This review is very well written, it summarizes the main evidence supporting the phylosymbioisis concept across diverse species and with specific examples on how phylosymbioisis shapes host-microbiome associations and fitness across living systems. It also summarizes methodological steps to characterize phylosymbiosis.

I only have 2 suggestions on additional minor additions (sections) that would make a more complete review.

1. How are current meta-OMIC techniques to characterize the microbiome and that denote functional rather than taxonomic modules (metagenomics- metabolomics) fit in the phylosymbioisis concept? Most analyses of these type have been based on taxonomic microbiome surveys. Also, would we expect microbiome functionality (genes/pathways/metabolites) to be less constrained by host phylogeny than microbiome taxa/phylogeny?

2. Along these lines, a section with specific examples in which phylosymbiosis has not been supported or in which host phylogenetic patterns does not support congruent microbiomes could be important. Specifically, the importance of shared dietary niche in different (less related) hosts, in shaping or rejecting phylosymbiosis could be discussed.

Author's Response to Decision Letter for (RSPB-2019-1539.R0)

See Appendix A.

RSPB-2019-2900.R0

Review form: Reviewer 1

Recommendation
Accept with minor revision (please list in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Good

General interest: Is the paper of sufficient general interest?
Good

Quality of the paper: Is the overall quality of the paper suitable?
Good

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
The revised version of this review is substantially improved. I think the authors did a good job adjusting the language to be more in line with an achievable scope. It now reads to me as a useful starting point for researchers intersted in the associated methods to identify relevant literature,

rather than seeking to be a comprehensive set of recommendations. I have a few small remaining quibbles and suggestions.

109-110: On first impression, this example might not seem like phylosymbiosis: here, experimental manipulation of a community induces genomic changes in separate experimental groups of host species. To someone unfamiliar with phylosymbiosis, this could seem backwards. If you have the space, explicitly calling out the fact that the pattern or correlation could arise this way, too, might help solidify the concept for the reader.

120: 'evolution-informed' or 'evolutionarily informed' rather than evolutionary-informed

120-121: I don't think you can really have 'evolutionary selective pressures for phylosymbiosis' -- you can have selective pressures that result in phylosymbiosis.

161-163: Rather than cast aspersions on 'early misconceptions' and 'strictly narrow presumptions', I'd suggest just straightforwardly stating that phylosymbiosis does not require or imply these things. This will likely be more clear to a reader unfamiliar with the existing literature.

It might also be worth eliminating the first line, which accounts for almost 15% of this entire section now.

228-231: If word count limitation is indeed such a problem for this review, is restating the formula for the Bray-Curtis metric really worthwhile?

232-234: UniFrac is really using the phylogenetic distances between microbes to estimate a phylogenetic distance between communities. It might also be worth specifying in 234 that the use of a phylogenetic tree *of microbes* is necessitated.

249-250: it seems like an explicit definition is in order here. E.g. "Microbiome distinguishability, or the characteristic of being able to statistically significantly differentiate the communities of host lineages under evaluation, is a prerequisite…"

Is this something that should be done by researchers initially before subsequently performing tests of phylosymbiosis?

275: please note here which tree (host or microbe) is used here.

285: If use of both metrics is 'strongly recommended' it seems reasonable to include some sort of justification.

287: Again, because the possibility of affecting false positive rate by including conspecifics isn't specifically mentioned earlier, I think it is important to include here at least some mention of how randomization is done. Simply randomizing tip labels without maintaining conspecific relationships can result in a 'significant' result even in the absence of phylosymbiosis. This is such an important and common stumbling block that I think it is important to address directly, rather than through the oblique comment about 'more study being necessary' in the earlier section.

363-364: Wouldn't at least some of these by definition no longer be phylosymbiosis? Aren't linkages between microbial diversity, function, and environmental factors... just plain old microbial ecology?

398-399: I'm not sure what is meant by "when host phylogeny is examined given host identity."

437: I don't agree with the shorthand here connecting 'universality' and falsifiability. The universality of a pattern has nothing to do with its being testable. It could be the case that every

single microbiome on the planet was correlated with host phylogeny, or none of them—the pattern would still be testable.

Rather, the fact that we see differences in the strength and significance of phylosymbiosis across different systems implies that there are underlying differences in the strength or type of mechanisms generating that signal, which could in turn help us design studies that usefully elucidate those differences. The non-universality of the pattern is thus important, not for its testability, but for its utility.

442-444: I don't know that this comparison is useful, especially without any attempt to define dysbiosis, explore how it's been used, and specifically detail its deficiencies as a term vis a vis phylosymbiosis.

451: if phyosymbiosis is a pattern and not a mechanism, it can't accelerate or decelerate -- or do -- anything. It could be associated with accelerated or decelerated speciation.

Decision letter (RSPB-2019-2900.R0)

15-Jan-2020

Dear Dr Bordenstein:

Your manuscript has now been peer reviewed and the reviews have been assessed by an Associate Editor. The reviewer's comments (not including confidential comments to the Editor) and the comments from the Associate Editor are included at the end of this email for your reference. As you will see, the reviewers and the Editors are positive but have raised some issues that we would like you to address in a revision.

We do not allow multiple rounds of revision so we urge you to make every effort to fully address all of the comments at this stage. If deemed necessary by the Associate Editor, your manuscript will be sent back to one or more of the original reviewers for assessment. If the original reviewers are not available we may invite new reviewers. Please note that we cannot guarantee eventual acceptance of your manuscript at this stage.

To submit your revision please log into http://mc.manuscriptcentral.com/prsb and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions", click on "Create a Revision". Your manuscript number has been appended to denote a revision.

When submitting your revision please upload a file under "Response to Referees" in the "File Upload" section. This should document, point by point, how you have responded to the reviewers' and Editors' comments, and the adjustments you have made to the manuscript. We require a copy of the manuscript with revisions made since the previous version marked as 'tracked changes' to be included in the 'response to referees' document.

Your main manuscript should be submitted as a text file (doc, txt, rtf or tex), not a PDF. Your figures should be submitted as separate files and not included within the main manuscript file.

When revising your manuscript you should also ensure that it adheres to our editorial policies (https://royalsociety.org/journals/ethics-policies/). You should pay particular attention to the following:

Research ethics:
If your study contains research on humans please ensure that you detail in the methods section whether you obtained ethical approval from your local research ethics committee and gained informed consent to participate from each of the participants.

Use of animals and field studies:
If your study uses animals please include details in the methods section of any approval and licences given to carry out the study and include full details of how animal welfare standards were ensured. Field studies should be conducted in accordance with local legislation; please include details of the appropriate permission and licences that you obtained to carry out the field work.

Data accessibility and data citation:
It is a condition of publication that you make available the data and research materials supporting the results in the article. Datasets should be deposited in an appropriate publicly available repository and details of the associated accession number, link or DOI to the datasets must be included in the Data Accessibility section of the article (https://royalsociety.org/journals/ethics-policies/data-sharing-mining/). Reference(s) to datasets should also be included in the reference list of the article with DOIs (where available).

In order to ensure effective and robust dissemination and appropriate credit to authors the dataset(s) used should also be fully cited and listed in the references.

If you wish to submit your data to Dryad (http://datadryad.org/) and have not already done so you can submit your data via this link
http://datadryad.org/submit?journalID=RSPB&manu=(Document not available), which will take you to your unique entry in the Dryad repository.

If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link.

For more information please see our open data policy http://royalsocietypublishing.org/data-sharing.

Electronic supplementary material:
All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI. Please try to submit all supplementary material as a single file.

Online supplementary material will also carry the title and description provided during submission, so please ensure these are accurate and informative. Note that the Royal Society will not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details (authors, title, journal name, article DOI). Your article DOI will be 10.1098/rspb.[paper ID in form xxxx.xxxx e.g. 10.1098/rspb.2016.0049].

Please submit a copy of your revised paper within three weeks. If we do not hear from you within this time your manuscript will be rejected. If you are unable to meet this deadline please let us know as soon as possible, as we may be able to grant a short extension.

Thank you for submitting your manuscript to Proceedings B; we look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Best wishes,
Professor Hans Heesterbeek

mailto: proceedingsb@royalsociety.org

Associate Editor Board Member
Comments to Author:
Thank you for your revised contribution to the special issue. The manuscript has been reviewed by myself and one of the original reviewers, and we both find the work both greatly improved and suitable for publication pending minor revisions. We think the paper will make a nice contribution to the field as a primer for those entering this fast moving topic.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s).
The revised version of this review is substantially improved. I think the authors did a good job adjusting the language to be more in line with an achievable scope. It now reads to me as a useful starting point for researchers intersted in the associated methods to identify relevant literature, rather than seeking to be a comprehensive set of recommendations. I have a few small remaining quibbles and suggestions.

109-110: On first impression, this example might not seem like phylosymbiosis: here, experimental manipulation of a community induces genomic changes in separate experimental groups of host species. To someone unfamiliar with phylosymbiosis, this could seem backwards. If you have the space, explicitly calling out the fact that the pattern or correlation could arise this way, too, might help solidify the concept for the reader.

120: ‘evolution-informed’ or ‘evolutionarily informed' rather than evolutionary-informed

120-121: I don't think you can really have 'evolutionary selective pressures for phylosymbiosis' -- you can have selective pressures that result in phylosymbiosis.

161-163: Rather than cast aspersions on ‘early misconceptions' and 'strictly narrow presumptions’, I'd suggest just straightforwardly stating that phylosymbiosis does not require or imply these things. This will likely be more clear to a reader unfamiliar with the existing literature.

It might also be worth eliminating the first line, which accounts for almost 15% of this entire section now.

228-231: If word count limitation is indeed such a problem for this review, is restating the formula for the Bray-Curtis metric really worthwhile?

232-234: UniFrac is really using the phylogenetic distances between microbes to estimate a phylogenetic distance between communities. It might also be worth specifying in 234 that the use of a phylogenetic tree *of microbes* is necessitated.

249-250: it seems like an explicit definition is in order here. E.g. "Microbiome distinguishability, or the characteristic of being able to statistically significantly differentiate the communities of host lineages under evaluation, is a prerequisite…”

Is this something that should be done by researchers initially before subsequently performing tests of phylosymbiosis?

275: please note here which tree (host or microbe) is used here.

285: If use of both metrics is 'strongly recommended' it seems reasonable to include some sort of justification.

287: Again, because the possibility of affecting false positive rate by including conspecifics isn't specifically mentioned earlier, I think it is important to include here at least some mention of how randomization is done. Simply randomizing tip labels without maintaining conspecific relationships can result in a 'significant' result even in the absence of phylosymbiosis. This is such an important and common stumbling block that I think it is important to address directly, rather than through the oblique comment about 'more study being necessary' in the earlier section.

363-364: Wouldn't at least some of these by definition no longer be phylosymbiosis? Aren't linkages between microbial diversity, function, and environmental factors... just plain old microbial ecology?

398-399: I'm not sure what is meant by "when host phylogeny is examined given host identity."

437: I don't agree with the shorthand here connecting 'universality' and falsifiability. The universality of a pattern has nothing to do with its being testable. It could be the case that every single microbiome on the planet was correlated with host phylogeny, or none of them—the pattern would still be testable.

Rather, the fact that we see differences in the strength and significance of phylosymbiosis across different systems implies that there are underlying differences in the strength or type of mechanisms generating that signal, which could in turn help us design studies that usefully elucidate those differences. The non-universality of the pattern is thus important, not for its testability, but for its utility.

442-444: I don't know that this comparison is useful, especially without any attempt to define dysbiosis, explore how it's been used, and specifically detail its deficiencies as a term vis a vis phylosymbiosis.

451: if phyosymbiosis is a pattern and not a mechanism, it can't accelerate or decelerate -- or do -- anything. It could be associated with accelerated or decelerated speciation.

Author's Response to Decision Letter for (RSPB-2019-2900.R0)

See Appendix B.

Decision letter (RSPB-2019-2900.R1)

07-Feb-2020

Dear Dr Bordenstein

I am pleased to inform you that your manuscript entitled "An Introduction to Phylosymbiosis" has been accepted for publication in Proceedings B.

You can expect to receive a proof of your article from our Production office in due course, please check your spam filter if you do not receive it. PLEASE NOTE: you will be given the exact page

length of your paper which may be different from the estimation from Editorial and you may be asked to reduce your paper if it goes over the 10 page limit.

If you are likely to be away from e-mail contact please let us know. Due to rapid publication and an extremely tight schedule, if comments are not received, we may publish the paper as it stands.

If you have any queries regarding the production of your final article or the publication date please contact procb_proofs@royalsociety.org

Open Access
You are invited to opt for Open Access, making your freely available to all as soon as it is ready for publication under a CCBY licence. Our article processing charge for Open Access is £1700. Corresponding authors from member institutions (http://royalsocietypublishing.org/site/librarians/allmembers.xhtml) receive a 25% discount to these charges. For more information please visit http://royalsocietypublishing.org/open-access.

Your article has been estimated as being 10 pages long. Our Production Office will be able to confirm the exact length at proof stage.

Paper charges
An e-mail request for payment of any related charges will be sent out after proof stage (within approximately 2-6 weeks). The preferred payment method is by credit card; however, other payment options are available

Electronic supplementary material:
All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

Thank you for your fine contribution. On behalf of the Editors of the Proceedings B, we look forward to your continued contributions to the Journal.

Sincerely,
Professor Hans Heesterbeek
Editor, Proceedings B
mailto: proceedingsb@royalsociety.org

Associate Editor:
Comments to Author:
Thank you for your revisions, which we feel have addressed all remaining concerns, We look forward to including the manuscript in the upcoming special issue.

Appendix A

VANDERBILT UNIVERSITY

Seth R. Bordenstein
Departments of Biological Sciences &
Pathology, Microbiology, and Immunology
Director, Vanderbilt Microbiome Initiative
Associate Director, Vanderbilt Institute for Infection, Immunology, and Inflammation

December 11, 2019

Dear Dr. Hans Heesterbeek and Colleagues,

Thank you for the detailed and constructive feedback on our manuscript, RSPB-2019-1539. We appreciate the opportunity to improve this work and hope the changes tracked in the main text and re-evaluation provides necessary clarity for a final favorable decision.

Seth R. Bordenstein
Vanderbilt University
Phone: 615.322.9087
Email: s.bordenstein@vanderbilt.edu

Responses to Reviewers

Referee: 1

Lim and Bordenstein have taken on an important challenge with this review, and one that I think is timely: providing a touchstone article precisely defining the concept of phylosymbiosis, reviewing the cases in which it has been documented, justifying the utility of the concept, and outlining how it can be used to guide research into the future. Use of the term is not without some controversy in the field. I find it to be a useful concept, but frequently find myself having to defend it to more skeptical colleagues. A cleary and carefully written article that methodically laid out precisely what constituted phylosymbiosis and why it is a useful term would make that job much easier, and I think would do a lot to improve research in host-associated microbiomes generally.

This is not that, at least not yet. I found the writing to be far too vague to be useful throughout much of the manuscript. For some of my phylosymbioskeptical colleagues, I worry that the manuscript as presented would only harden their conviction that the term is of little use. There is too much assertion *that* phylosymbiosis is useful, and far too little evidence for how and why. I think the paper would benefit tremendously from some very carefully considered and spelled-out thought experiments, perhaps even employing simulated or toy data, showing exactly how a specific measurement (the correlation between microbiome dissimilarity and host phylogeny) could be used to interpret or guide investigation of these complex systems.

One of the most problematic aspects of phylosymbiosis, for some, is that there are many possible underlying mechanisms to generate such a pattern, as well as many possible ways to measure it. This needs to be tackled more directly. Some of my favorite papers on the subject, such as Groussin et al, have used this to their advantage by making explicit connections between the different processes that might be underlying phylosymbiotic patterns and their predicted effects on different ways of measuring these patterns. I understand that in trying to write an "introduction," the authors might be wary of getting too far into the weeds of details. But absent that kind of specificity, I worry that the authors' advice to report measurements of phylosymbiosis from across a wide range of parameter values will sound too much like throwing as much as possible at a wall to see what sticks. I think in its best possible realization, this article would engage very carefully with the implications of these details to help convey to a reader why and how a concept of phylosymbiosis has utility.

We thank the reviewer for these suggestions. Indeed, as the reviewer suspected, we originally approached this "Introduction to Phylosymbiosis" for newcomers and beginners as there is a palpable need to bring the non-expert community into the basics of phylosymbiosis. That was our primary goal, but we realize that the reviewers, who are likely experts in the field, would like more detail. With that in mind, we revised major sections of the manuscript to more explicitly highlight the specificities and evidence of phylosymbiosis, while taking into account the severe journal word limit (7,500 words including references with two figures) that will continue to pose problems in introducing phylosymbiosis to beginners while adding further detail for experts in the review. Please see our more complete replies below.

Below are specific comments. I apologize for the very long list, which I know will seem onerous to address. Please be assured that they come from a sincere desire to improve the manuscript, and a great desire to see it published. I will restate that I think such a paper could be of great value to the field. Good luck!

51-54: This doesn't seem clear or specific enough of a definition to me -- wouldn't 'associations between two or more organisms' without temporal or categorical restriction encompass *any* association between organisms? Is that restrictive enough to be considered a useful and clear definition?

Thank you for the comment. Phylosymbiosis distinguishes itself from non-phylosymbiosis by characterizing a significant degree of association between host phylogenetic and microbiome community relationships. It is not universal, and thus we do think it provides a testable hypothesis, reflects the variation likely to be seen in nature, and is otherwise assumption-free and amenable to explanation by a variety of eco-evolutionary mechanisms. We also think the rising number of publications evaluating phylosymbiosis versus non-phylosymbiosis adds evidence to its utility. The determination of whether phylosymbiosis is present or not is a first step preceding further investigations into mechanistic details, such as the nature of species-species associations and the type(s) of ecological and evolutionary processes shaping phylosymbiosis. Given the growing acceptance of the term and its varied tests (topological congruency, Mantel, etc.), phylosymbiosis does seem far clearer and relatively more specific in the literature than other widely-used terms such as dysbiosis or microbiome versus microbiota. If the microbiome field will have general trends to test in new systems, phylosymbiosis seems well-poised for this circumstance.

80: The use of the phrasing "a host phylogenetic effect on" seems problematic to me, as it implies a causal relationship from the phylogeny to the microbial community. I understand that in some contexts "a phylogenetic effect" can be used as statistical shorthand when talking about correlations, but given that this is meant to be an explanatory piece, I worry about implying causation so early on.

We fully agree with the reviewer and thus replaced the phrase with "a significant association between host phylogenetic relationships and host-associated microbial community relationships" (lines 81-82 in revised manuscript).

95-127: This section feels a little disjointed—the examples read a little like a list, with little to provide context or a narrative for the reader. These three paragraphs are followed by a final paragraph exploring potential adaptive significance of host-specific microbiomes. Is this part of the "history of phylosymiosis?" Why?

Yes, it was originally supposed to be part of a section on the "history of phylosymbiosis". In line with suggestions from both reviewers, we revised the manuscript to more seamlessly incorporate results and predictions from previous studies.

101: I would drop "randomized and non-randomized" here and simply say "comparisons of beta diversity-derived clusters with mammalian phylogenies...". The randomization is how the strength of the comparison was assessed, but without additional context it might seem confusing.

Great point, and we apologize for the confusion. In the process of revising the first half of the manuscript, these statements have been condensed and moved elsewhere (lines 86-91 in revised manuscript). The phrase "randomized and non-randomized" has been removed.

116: Is "functional genetic" the best term here? Some readers, particularly from more ecological or natural history backgrounds, may not have an immediate sense of what is implied by "functional genetics" in this context. Why not simply say "The first study linking phylosymbiotic patterns to the function of specific host genes"?

Good point. We replaced the phrase "functional genetic" with "The first study linking phylosymbiotic patterns to the function of specific host genes" (lines 102-103 in revised manuscript).

132: "Microbiome studies have demonstrated that resident microbes can preferentially colonize some host taxa, likely through host-specific... colonization (20)." This seems like a truism as written?

We agree and removed this statement because it is not crucial to the manuscript.

134: "through non-immune and immune factors" would seem to just about cover the possible types of factors... is it necessary to specify?

We agree that this specification is not necessary. In revising the first half of the manuscript, we removed the statement containing the phrase.

135: What is "functional phylosymbiosis"?

We now explicitly refer to "functional phylosymbiosis" as "when host and/or microbial phenotypes impact or are impacted by phylosymbiotic associations" (lines 123-124 in revised manuscript).

136-137: Why specifically might hybridization between closely related host species disrupt host-microbiome associations?

To clarify this point, we added the following text to lines 141-149 of the revised manuscript:

> "Hybridization between host species causes host-microbiome mismatches since combining independently-evolved host genotypes in a hybrid may cause a breakdown in either microbial colonization preferences for certain hosts or host control of the microbiome. As demonstrated in *Nasonia* (9), house mice (26), and whitefish (27), hybrids have an altered microbiome relative to the parental microbiome, suggesting a reduced capacity for hosts to regulate their microbiomes as well as an increased capacity for pathogenic microbes to bloom. These breakdowns in host-microbiome interactions can associate with maladaptive phenotypes in hybrids including immune dysfunction, pathology, inviability, and sterility (9,26) that can reduce interbreeding between species or populations."

143-144: Is this really a fair generalization of the results of the Nasonia experiments cited? I would agree that the experiment demonstrates how microbes can be an integral component of hybrid sterility and post-zygotic isolation, and that this in turn can help explain how phylosymbiotic patterns might arise. But is it really the phylosymbiotic-ness of the community that's leading to adverse consequences?

We appreciate the point and revised this phrasing in lines 153-157 of the revised manuscript:

> "Collectively, the results from interspecific microbiome transplant experiments and host hybridization studies illustrate that host-microbiome interactions across host species can have important functional consequences that impact evolutionary outcomes within and between species, including wedging host populations into species."

165: This is just a linguistic quibble, but can host ancestors split into different but _un_related genera?

Thanks. This statement has been removed as part of the manuscript revision.

175-178: Earlier, you state (and I think accurately) that "a positive association between host phylogenetic and microbial community relationships does not a priori imply a shared and ancestral evolutionary history... it may or may not reflect long-term associations or co-adaptations." Here, though, you state that "specialized host-microbe associations indicate that hosts are adapted to their native microbiomes." These statements seem in some conflict!
Also, I don't see how these examples indicate that the given phylosymbiotic patterns arose via natural selection and not neutral evolution. Neutral drift can absolutely lead to phenotypic consequences even in the absence of selection.

We apologize for the confusion. While the former statement applies to phylosymbiosis in general, we intended the word "specialized" here to apply to the aforementioned case examples where microbiomes impact host fitness. That was not clear, so the statement has been revised:

> "Collectively, phylosymbiotic associations that impact host fitness support the premise that hosts are adapted to their native microbiomes rather than non-native microbiomes, although more studies are needed to confirm these associations and effects in captive and wild host populations." (lines 136-139 in revised manuscript)

193: What are "phylosymbiosis outcomes"?

"Phylosymbiosis outcomes" means the same thing as "phylosymbiosis", and we have removed the word "outcome" in the revised manuscript to avoid confusion.

195-196: This statement seems like it could be expanded. How is 'stochastically' being used here? One could imagine stochastic processes–such as geographically isolated host populations exposed to different microbial metacommunities–that would result in a pattern of phylosymbiosis.

To me, it would be more useful here to explicitly consider a couple different scenarios that might produce a pattern of phylosymbiosis, rather than simply asserting a statement like this.

We originally mentioned the role of physical barriers in the "Future Directions" section of the original manuscript (lines 553-556): "A study on predator-prey pairs of wild mammalian species showed that divergence in gut microbial beta diversity can be accelerated by physical barriers and, inversely, reduced by predator-prey interactions between host species."

Nevertheless, we agree with the reviewer that geographic isolation, among others, could be one of the causes. In revising the first half of the manuscript, we more carefully outlined possible scenarios and explanations for phylosymbiosis, including geographically isolated host populations exposed to different microbial communities:

> "Phylosymbiosis may arise from stochastic and/or deterministic evolutionary and ecological forces. For example, stochastic effects include dispersal fluctuations in microbial communities (ecological drift) or shifts in host geographic ranges (14)." (lines 96-98 in revised manuscript)

However, due to the journal's severe length restrictions, we are unfortunately unable to provide further details on how stochastic processes contribute to phylosymbiosis, and this is not the main focus of the manuscript. Other scenarios that contribute to phylosymbiosis are mentioned in lines 98-110 in the revised manuscript:

204-205: Where does the recommendation for an N of 4 come from? 24 is not a tremendous number of possible permutations, meaning that any analysis with just 4 taxa is going to be extremely underpowered.

The recommendation of N=4 comes from our Robinson-Foulds and Matching Cluster analyses of congruency (Brooks et al, 2016) in which rooted trees with 4 taxa can yield a minimum p-value of 0.042. That being said, we agree with the reviewer, and removed the recommendation in the revised manuscript. We revised the section on "Host input data" (now "Host taxa and input data") where we 1) briefly discuss the importance of statistical power; 2) recommend that power and effect size analyses be performed prior to data collection; and 3) recommend that the experimental design of previously successful studies be used as reference for phylosymbiosis detection:

> "*Host taxa and input data*. Because phylosymbiosis detection involves the collection of replicated samples across multiple taxa, both optimization of statistical sensitivity (28) and specificity (18) as well as minimization of sequencing batch effects are crucial for differentiating between noise and signal. Although our 2016 study showed that rooted trees with four *Nasonia* species are sufficient to detect phylosymbiosis within the clade (23), we suggest the use of appropriate power and effect size analyses (reviewed in (29) for microbiome data) to determine sufficient replicates and taxa for the optimization of statistical power (28). Sampling multiple individuals per species will help resolve noise from signal in microbial community relationships, but further study is required on how replicates of inter- and intra-species samples are best utilized in studying phylosymbiosis across host clades that can vary in divergence times. If available, experimental designs of successful phylosymbiosis studies with similar sample types can also be adapted accordingly (30)." (lines 180-192 in revised manuscript).

218: This section makes no mention of the quandary of how to represent multiple individuals from the same species in an assessment of phylosymbiosis, which was noted by Mazel et al to have the potential to vastly inflate false positives.

We revised the section on "Host input data" (now "Host taxa and input data") to address this – see response above.

220: Similar to above, where does a recommendation of ten samples come from? What is the statistical estimate whose confidence is being maximized here? Increasing the number of samples per species will improve the estimate of that taxon's microbiome, but has no bearing on the power of the estimate of phylosymbiosis itself, which is ultimately constrained by the number of taxa assessed.

We removed the recommendation of ten samples which was based on some unpublished analyses in the lab. We also revised the section on "Host input data" (now "Host taxa and input data") to address the author's concerns – see response above.

238-239: Why should quantitative metrics be preferred? For that matter, why should *any* type of metric be preferred? They are all measurements of community dissimilarity that reflect different aspects of that dissimilarity. Quantitative metrics, while less sensitive to rare taxa, can be overwhelmed by variation in a few very abundant ones. Which is appropriate for a particular study?

We appreciate the reviewer's feedback and removed the emphasis on quantitative metrics. Please see the revised text (lines 236-247 in revised manuscript and below) for the changes:

> "Because beta diversity metrics reflect different aspects of dissimilarity, the choice of metric is study specific and depends partly on the microbial composition and evolutionary history of the lineages studied. Binary metrics based on presence/absence are more sensitive to variations in rare taxa and were implemented to study host specificity of sponge microbiomes, where rare taxa comprised more than 90% of distinct OTUs (50). Binary metrics may also be sensitive to recent microbial diversification because recently diverged OTUs/ASVs will exert the same effect as OTUs/ASVs with a longer divergence history (39). In contrast, quantitative metrics are more sensitive to variations in abundant taxa. Besides taxonomy-based phylosymbiosis studies (23,51-53), quantitative metrics have also been applied to metagenomics data (42,43). Metrics that consider phylogenetic relationships between OTUs, such as UniFrac distances, (54) are applied in many other phylosymbiosis studies, including bats (55), corals (20), and mammals (4,43).

247-249: How should researchers interpret the inevitable differences among these parameter combinations? This also raises the specter of multiple testing and specificity, which should also be discussed in the context of recommended size of the taxon sample. How should we interpret when one of twelve tested parameter combinations reveals phylosymbiosis?

This is an excellent point without a simple, short answer, and we have added a brief section on parameter selection in the revised manuscript to address this (lines 325-334 in revised manuscript):

> *"Parameter selection.* Phylosymbiosis detection involves selection of various parameters, such as OTU identity cutoff, beta diversity metric, clustering method, and congruency test, each with their strengths and limitations that will vary with study design and questions. Although various parameter combinations can be tested and compared simultaneously (39), in the case when only a few of all possible parameter combinations detect phylosymbiosis, we recommend cautious interpretation of results with respect to the chosen parameters. If available, results should also be compared to those from previous phylosymbiosis studies with similar sample types using the same parameter combinations. Experimental replication is also necessary to confirm phylosymbiosis, especially when it is not consistently detected."

255: How does distinguishability relate to phylosymbiosis? Is distinguishability a prerequisite for phylosymbiosis?

We note the reviewer's comment and have clarified this in lines 249-250 in the revised manuscript:

> "Microbiome distinguishability, representative of microbial beta diversity differences between host lineages under evaluation, is a prerequisite for phylosymbiosis (20,23,51-53)."

280: Microbiome similarity dendrograms are also somewhat arbitrary (why use UPGMA clustering, instead of one of the many alternatives?), and especially when the number of samples analyzed increases, can have very unstable topologies. How is this to be interpreted in the context of phylosymbiosis?

We appreciate the reviewer's points and have modified the section (lines 263-277 in revised manuscript and below) where we explicitly mentioned the availability of multiple clustering methods and briefly compared a few of them.

We are, however, unable to describe clustering approaches beyond this detail due to journal length restrictions and the broad intent to make this article an introduction for non-experts. Instead, we referred readers to Legendre and Legendre's book on Numerical Ecology (highlighted in bold below) for in-depth descriptions and comparisons of conventional clustering methods.

> "To generate a hierarchical dendrogram, several agglomerative hierarchical clustering methods **(reviewed in (56))** can cluster microbial beta diversity distances. The most commonly used method, unweighted pair group method with arithmetic mean (UPGMA), performs pairwise sample clustering from their average dissimilarity values and gives all samples equal weights (60). Compared to linkage clustering approaches, UPGMA prioritizes relationships among groups over individual samples (56). By assigning equal weights to all samples, UPGMA assumes that samples in each group are representative of groups in the larger reference population (56). As such, it may be sensitive to sample sizes and may generate unstable topologies with imbalanced data where some groups are oversampled while some are undersampled. Newer clustering methods, such as the phylogenetically-aware squash clustering method, directly compute distances between samples (rather than differences between beta diversity distances) based on their positions on a phylogenetic tree (61). In general, the effects of clustering methods on phylosymbiosis detection require further study.

296-298: I don't quite understand this sentence. Is it meant to be specific to matrix correlation methods? Also to topology-based methods?

These lines originally referred to both matrix correlation and topology-based methods. This statement has been removed as part of the manuscript revision.

Additionally, Mazel et al found matrix correlation methods to have significant advantages over topology-based approaches using their simulated data benchmarks. This seems relevant to mention here, no?

Good point. We added Mazel et al.'s findings to lines 312-323 of the manuscript and note caveats of this analysis:

> "A simulation analysis suggested that the Mantel test has higher sensitivity and power than the Robinson-Foulds metric when phylosymbiosis is based on the assumption of microbial preferences for a host trait (19). The practical relevance of this conclusion is not clear because phylosymbiosis will arise from reasons other than microbial colonization preferences, such as host preferences, neutral processes, and microbe-microbe interactions. Moreover, the performance between Mantel test and the more sensitive topology-based Matching Cluster distance was not evaluated in this simulation, and such comparisons are likely to yield different insights. Systematic benchmarking of type I and II error rates of phylosymbiosis measurement methods across various possible scenarios will aid experimental design and result interpretation. As such, research opportunities for the development and implementation of improved phylosymbiosis detection methods are ample.

342: "gender" should be replaced with "sex" in this context.

Thanks! This statement has been condensed to:

> "In *Drosophila* flies, phylosymbiosis patterns are either weakly supported (23) or not detected (80) in both lab strains and wild populations." (lines 372-374 in revised manuscript).

344-347: I'm having trouble with this sentence. What is a cluster of metagenomes? And is "microbial taxonomic order" meant to be "host taxonomic order"? (It would be a truism that OTUs would vary with microbial taxonomy, no?) Why would covariance of bacterial relative abundance and host genes imply a functional relationship? It seems equally possible that this sort of correlation could be indirect.

We apologize for the structure and lack of clarity of these lines. This sentence has been condensed (see response to comment above).

377-379: Is correlation of microbial alpha diversity with host phylogeny considered a facet of phylosymbiosis? If so, it should probably be explored more explicitly early on.

It is not crucial to understanding 'what phylosymbiosis is' and because of the word length restriction, we have removed these lines in the revised manuscript as we add other details requested by the reviewers. Thanks for pointing this out.

379: I worry that referencing per-lineage cophylogeny here will tend to muddle the distinction of 'what phylosymbiosis is' to a new reader. Having examples of cophylogeny or correlations with alpha diversity mixed in with these examples might be counterproductive -- would it be worth instead compartmentalizing them in a brief additional section on other phylogenetic correlations?

We thank the reviewer for the feedback. Because this topic is not crucial to understanding 'what phylosymbiosis is' and because of the word length restriction, we removed these lines in the revised manuscript to accommodate adding other details requested by the reviewers.

383: 'bacterial bacteriomes' is redundant

It definitely is. Thank you for pointing this out. We have removed the word "bacterial" and revised the sentence to (replacement for 'bacterial bacteriomes' in bold):

> "Phylosymbiosis and host dietary impacts also occur on the skin microbiomes of 44 fish species from the Western Indian Ocean (89), but do not exist on the **surface microbiomes** of sympatric kelp species (90)." (lines 402-404 in revised manuscript)

397-398: is it reasonable to discuss 'non-statistically significant phylosymbiotic correlations'?

We mentioned studies with 'non-statistically significant phylosymbiotic correlations' with the intention of presenting a balanced overview of both positive and negative phylosymbiosis examples. However, we agree that the level of detail for these studies in our original submission was not necessary and have shortened these lines in the manuscript:

> "Contrarily, qualitative incongruency between Brassicaceae host phylogeny and their root microbiomes has been observed (97), whereas non-statistically significant phylosymbiotic correlations have been reported in other plant microbiome studies (16,98)." (lines 417-420 in revised manuscript)

402-428: This section should be one of the highlights of this review—anyone skeptical of the utility of the term and concept is likely to skip to this heading to find out what IS the significance of phylosymbiosis. I'm worried that their conclusion might be "there is none." Most of what is presented are assertions *that*

phylosymbiosis is or provides something, rather than convincing examples or evidence for *why* and *how*.

We appreciate the reviewer's feedback and have revised this section ("Significance and future directions of phylosymbiosis") in lines 422-455 of the revised manuscript:

> "Microbiome research will continue to be revolutionized by the multi-omics era, where a deluge of data has enabled unprecedented insights into the extensive taxonomic, genetic, and functional composition of microbial communities and their associated hosts. Such large-scale accumulation of empirical and theoretical findings can potentiate the development of new hypotheses, unifying concepts, and frameworks across diverse host-microbiome systems. Indeed, the recurrence of phylosymbiosis across host systems lends itself to large comparative surveys across kingdoms of life that may uncover taxonomic range restrictions of phylosymbiosis as well as the environmental parameters (e.g., soil and water properties) and ecological interactions (e.g., diet and predator-prey relationships) that determine the boundaries of where and when phylosymbiosis occurs. If the microbiome field will have general trends to test in new systems, phylosymbiosis is well-poised for this circumstance.
>
> Phylosymbiosis distinguishes itself from non-phylosymbiosis by characterizing a significant degree of association between host phylogenetic and microbiome community relationships. It is not universal, and thus provides a testable hypothesis, reflects the variation likely to be seen in nature, and is amenable to explanation by mechanisms that require further investigation. The determination of whether phylosymbiosis is present or not is a first step preceding further investigations into mechanistic details, such as the nature of species-species associations and the type(s) of ecological and evolutionary genetic processes underpinning phylosymbiosis. Given the growing evidence for the pattern and increasingly sophisticated tools available to detect phylosymbiosis, phylosymbiosis is relatively clearer and more specific than other terms such as dysbiosis.
>
> Phylosymbiosis also engenders a holistic view of ecology and evolution in which hosts are communities or holobionts whose microbial members can contribute to genetic and phenotypic variation subject to natural selection. Several questions that have been conventionally overlooked include what are the microbial effects on host allele frequencies? Does host gene flow in natural populations impact microbiome variation and phylosymbiosis? Does phylosymbiosis accelerate or decelerate host speciation? What are the genetic and mechanistic factors that regulate phylosymbiosis, and how do these factors vary across populations or species? Collectively, studies determining the magnitude of ecological, evolutionary, and genetic forces in structuring phylosymbiosis is an important area of future research."

414-415: This section would benefit tremendously from some concrete examples. I'm having a hard time understanding how a pattern can provide eco-evolutionary predictions. This is especially true for something like phylosymbiosis, which is a pattern that can conceivably be generated by quite a large number of rather unrelated processes.

We have revised this section in lines 422-455 of the revised manuscript (see response to comment above).

417-419: As with the previous sentence, this one is quite difficult for me to meaningfully parse. (What is an ecological mechanism in this context? What would make such a mechanism intrinsic, and intrinsic to what?)

We have removed the line and revised this section in lines 422-455 of the revised manuscript (see response to comment above).

421-428: How, exactly, does phylosymbiosis contribute to this growing school of thought? Please be specific. What precisely can a measure of a pattern like phylosymbiosis do to help us disentangle these diverse interactions—to 'determine the magnitude of each of these traits'?

We have revised this section in lines 422-455 of the revised manuscript (see response to comment above).

451-452: "species conservation" seems to come a little out of left field here. Is this referring to the conservation of symbiont species across evolutionary time within lineages, or the conservation of threatened species?

We originally intended this to mean the conservation of threatened species. However, we agree with the reviewer that possible applications of phylosymbiosis in species conservation may cause confusion, especially since this is not further explained in the manuscript. Therefore, we have now removed this phrase in the revised manuscript.

458: "rules and themes": what would be an example of a "rule" pertaining to phylosymbiosis, and how would that differ from a "theme"? Given that phylosymbiosis is just a pattern, why exactly would a standardized workflow for its measurement lead directly to establishment of such "rules"?

We agree that the phrase "rules and themes" at the end of the manuscript is confusing and non-conclusive, and have now revised the statement (lines 460-462 in revised manuscript) accordingly:

> "As research in this area proliferates, a definition, conceptual framework, and workflow for assessing phylosymbiosis will facilitate identification of phylosymbiotic host-microbe interactions."

461-462: I think this concluding sentence will be useful in reassessing the rest of the paper. I completely agree that we need a much more thorough understanding of the mechanisms underlying host-microbe associations. I *also* agree that documentation of high-level patterns, like phylosymbiosis, can play a really useful role in the investigation of such processes. But much more work needs to be done in the piece to very carefully draw explicit connections from the observation of pattern to how, precisely, such an observation is likely to guide investigation of process.

We thank the reviewer for the feedback and revised the rest of the paper keeping this comment and other comments from all reviewers in mind. We hope that the revised manuscript addresses the reviewer's concerns satisfactorily.

Referee: 2

The authors present a review of host-microbiome "co-phylogeny" studies (i.e., studies comparing the phylogenies of hosts and with the phylogenies of their microbiota) within a framework informed by the concept of "phylosymbiosis". The phylosymbiosis framework is a very useful one for host-symbiont studies, Unfortunately, the paper (and probably the concept) needs some clarification and revisions before the framework is ready to be shared with the community. The authors promise presentation of a framework for studies of phylosymbiosis, with a review that "serves as a gateway to experimental, conceptual, and quantitative themes". As it stands, this paper falls short of this goal. In particular it fails to serve as a standalone description of the phylosybiosis framework: one has to have read, in depth, the Brooks et al.

2017 paper referenced many times in this work, to actually even know the statistics as well as predictions of "phylosybiosis", and having done that, it is not clear to me how much additional insight this paper contributes to the discussion.

Most of the problems lie in the first half the paper, where the phylosymbiosis framework is (purportedly) presented along with various methods for assessing phylosybiosis. This presentation is woefully inadequate and very disappointing, lacking in conceptual and statistical detail. Most of the methods are presented simply by name, with no value-added interpretation, explanation, or insight into their mechanics. Beyond the relatively narrow scope of statistical methodologies, however, in general the presentation of the phylosymbiosis framework is extremely disappointing. It is insufficient in detail to really understand the framework, and there is no additional insight or interpretation provided to make it worthwhile reading this paper after having read some of the primary works it itself references (such as the Brooks et al. 2017 paper).

We thank the reviewer for the feedback. We originally wrote a long and comprehensive review with technical details of statistical methods, terms, and relevant citations. Unfortunately, we had to remove these due to severe length restrictions by the journal. We asked the journal to relax the page restrictions, but the journal editors declined. We carefully considered comments from all the reviewers and have added some of these technical details into the manuscript.

- [LL 78-93] Here the authors qualitatively define "phylosymbiosis". The authors define phylosymbiosis as "microbial community relationships that recapitulate the history of the host". The authors then go on to say that phylosymbiosis is a "pattern observed in a snapshot in time and space, and may or may not reflect long-term associations or co-adaptations that can be subsequently evaluated by empirical analyses". Many issues with this paper could have been avoided if the authors went on to recapitulate the major points, predictions, and quantitative concepts from the Brooks et al., 2017 paper here as well, to clearly and explicitly define the phylosymbiosis framework.

We thank the reviewer for the suggestion and have revised the first half of the manuscript to focus on the definition and possible scenarios/explanations for phylosymbiosis (lines 50-169 in revised manuscript). The quantitative aspect of phylosymbiosis is detailed in the second half of the manuscript (lines 171-364).

- [LL. 81-83] The authors characterize "phylogeography" as "the study of evolutionary processes that shape geographic or ecological distributions of organisms".

Phylogeography is understood today as the study of the association between geographical and phylogenetic relationships, typically through the influence of the former on driving historical demographic processes (or events) that in turn structure the latter.

In this sense, the analogy with "phylosymbiosis" the authors is actually reverse in causality.

In phylosymbiosis, the "phylo" part (i.e., host's phylogenetic structure) is the casual component, informing the symbiote relationships, whereas in phylogeography, the "phylo" part is the dependent component, being informed by the geography.

This is a minor, and, in the larger picture, entirely inconsequential point, I know, but I mention it here nonetheless.

We thank the reviewer for the clarification. As mentioned by reviewer #1, phylosymbiosis does not immediately imply any causal relationship from the phylogeny to the microbial community or vice versa. Because we are expanding the manuscript to include more technical details as requested above, we removed the statements comparing phylosymbiosis to phylogeography since they are not crucial to the manuscript.

- [LL. 85-88] What is the difference between "host phylogenetic effect" (or definition of this) vs. "reciprocal evolutionary genetic changes" or "ancestral splitting of host and symbiotic lineages"? Without

a clear definition of the term to the contrary, I would understand "host phylogenetic effect" as possibly including the latter two.

With regards to the phrase “host phylogenetic effect”, a similar concern was raised by reviewer #1 and we replaced the phrase with “a significant association between host phylogenetic relationships and host-associated microbial community relationships” (lines 81-82 in revised manuscript) that does not imply causation.

- [LL. 101-103] "suggested a weak phylosymbiotic association"; as with "host phylogenetic effect" above, this term really should be defined before assertions or statements about it are made. In this case, the qualifier "weak" implies a quantitative aspect that needs to be explicitly described beforehand.

We agree with the reviewer and removed the phrase “weak phylosymbiotic association” from the first half of the manuscript to avoid confusion. The quantitative aspect of phylosymbiosis is detailed in the second half of the manuscript (lines 171-364).

- [LL. 135-135] "functional phylosymbiosis" -- the term is used here with no explicit definition. The reader might be able to infer its meaning from context, but (a) I'm not sure if two readers would agree; (b) in this case of this particular reader at least, the meaning I inferred is vague enough that I would not be able to explain it clearly to a third party; (c) in a work introducing this framework, this term should be explicitly defined and not left as an exercise to the reader.

We now define “functional phylosymbiosis” explicitly as “for when host and/or microbial phenotypes impact or are impacted by phylosymbiotic associations.” (lines 123-124 in revised manuscript).

- [LL 135-148] I am afraid that I could not follow this argument.
  - I was led to expect that I would be shown two ways to empirically test for "functional phylosymbiosis" (or what I inferred "functional symbiosis" to mean) by the statement "functional symbiosis can be evaluated empirically in at least two ways".

We apologize for the confusion. We removed the unnecessary quantifier “two ways” and revised the section accordingly to better describe how functional phylosymbiosis can be tested with interspecific microbiome transplant experiments and host hybridization studies (lines 126-157 in revised manuscript):

> “Interspecific microbiome transplant experiments are useful in elucidating functional phylosymbiosis. A large-scale phylosymbiosis investigation spanning 24 species across four laboratory-reared host clades (*Nasonia* wasps, *Drosophila* flies, mosquitoes, and *Peromyscus* deer mice) demonstrated that interspecific transplants of gut microbial communities between *Peromyscus* species decreased dry matter digestibility and increased food intake, while transplants between *Nasonia* species markedly lowered survival to adulthood by nearly half (23). In addition, interspecific microbiomes are more costly to *Nasonia* larval growth and pupation than intraspecific microbiomes (24). Similarly, reciprocal maternal symbiont transplants between two wild, sympatric *Ontophagus* dung beetle species caused developmental delay and elevated mortality in non-native hosts that persisted to the next generation (25). Collectively, phylosymbiotic associations that impact host fitness support the premise that hosts are adapted to their native microbiomes rather than non-native microbiomes, although more studies are needed to confirm these associations and effects in captive and wild host populations.
>
> Hybridization between host species causes host-microbiome mismatches since combining independently-evolved host genotypes in a hybrid may cause a breakdown in either microbial colonization preferences for certain hosts or host control of the microbiome. As demonstrated in

> *Nasonia* (9), house mice (26), and whitefish (27), hybrids have an altered microbiome relative to the parental microbiome, suggesting a reduced capacity for hosts to regulate their microbiomes as well as an increased capacity for pathogenic microbes to bloom. These breakdowns in host-microbiome interactions associate with maladaptive phenotypes in hybrids including immune dysfunction, pathology, inviability, and sterility (9,26) that can reduce interbreeding between species or populations. In *Nasonia*, lethality of hybrids between the older species pair was rescued by germ-free rearing and restored by feeding a 1:1 inoculum of the resident gut bacteria from parents to germ-free hybrids (9). In contrast, hybrids between a younger *Nasonia* species pair did not have an altered microbiome nor suffer functional costs. Collectively, the results from interspecific microbiome transplant experiments and host hybridization studies illustrate that host-microbiome interactions across host species can have important functional consequences that impact evolutionary outcomes within and between species, including wedging host populations into species."

- To follow through on this, the authors need to first clearly describe what the predictions of functional phylosymbiosis are, and then describe how empirical data can be tested for patterns conforming to these predictions. In addition, to be convincing, the authors need to demonstrate that there are no alternate explanations for these patterns or otherwise taken into account alternate explanations in their tests.

While phylosymbiosis is a general term to describe "microbial community relationships that recapitulate the history of the host" (line 80 in revised manuscript), functional phylosymbiosis is a general term "for when host and/or microbial phenotypes impact or are impacted by phylosymbiotic associations." (lines 123-124 in revised manuscript). We have described how functional phylosymbiosis can be studied in lines 126-157 of the revised manuscript (see response above).

- However, instead, for the first of their two evalations the authors present one possible outcome of hybridization (host-microbiome disruption in parental species) and go on to possible outcomes of this, in turn, in F2 generation in one particular system (Nasonia). The authors then conclude that "non-phylosymbiotic relationships can lead to adverse functional and evolutionary consequences over time".

We apologize for the confusion and have re-organized the section (lines 126-157 in revised manuscript) to improve the logic flow (see response above).

- First of all, the authors assertion that host-microbiom disruption in parental species as an outcome of hybridization needs to be demonstrated, at least conceptually or otherwise with reference to previous work.

In the first draft of the manuscript, we cited previous *Nasonia* work that demonstrated hybridization leads to microbiome disruption. The same also occurs for naturally hybridizing populations of house mice (see Wang *et al.*, *Nature Communications*, doi: 10.1038/ncomms7440). We hope lines 141 to 149 in the revised manuscript improves the clarity of this argument:

> "Hybridization between host species causes host-microbiome mismatches since combining independently-evolved host genotypes in a hybrid may cause a breakdown in either microbial colonization preferences for certain hosts or host control of the microbiome. As demonstrated in *Nasonia* (9), house mice (26), and whitefish (27), hybrids have an altered microbiome relative to the parental microbiome, suggesting a reduced capacity for hosts to regulate their microbiomes as well as an increased capacity for pathogenic microbes to bloom. These breakdowns in host-microbiome interactions associate with maladaptive phenotypes in hybrids including immune dysfunction, pathology, inviability, and sterility (9,26) that can reduce interbreeding between species or populations."

- More importantly, the authors have merely provided an interpretation of a pattern with reference to a "non-phylosymbiotic relationships" without actually showing how this is a prediction of functional phylosymbiosis let alone an exclusive prediction or serve as a satisfactory test for it.

This comment is similar to the reviewer's previous comment on the predictions of functional phylosymbiosis, which we have now more clearly addressed. Keeping in mind that functional phylosymbiosis is a term describing a scenario "when host and/or microbial phenotypes impact or are impacted by phylosymbiotic associations" (lines 123-124 in revised manuscript), the purpose of this section (lines 126-157 in revised manuscript) is to provide empirical examples of how disrupting phylosymbiotic associations can be used for the characterization of functional phylosymbiosis. We hope the revised manuscript makes this sufficiently clearer.

- The second part of this argument states that "microbiome transplant experiments between related species/lineages will lead to host fitness reductions ....". This may be the case, but it is unclear to me how this provides an assessment of "phylosymbiosis", at least as defined by the authors, i.e., "microbial community relationships that recapitulate the history of the host".

Thank you for the feedback. We have now clarified in lines 126-139 of the revised manuscript that the transplants can be used to evaluate functional phylosymbiosis.

> "Interspecific microbiome transplant experiments are useful in elucidating functional phylosymbiosis. A large-scale phylosymbiosis investigation spanning 24 species across four laboratory-reared host clades (*Nasonia* wasps, *Drosophila* flies, mosquitoes, and *Peromyscus* deer mice) demonstrated that interspecific transplants of gut microbial communities between *Peromyscus* species decreased dry matter digestibility and increased food intake, while transplants between *Nasonia* species markedly lowered survival to adulthood by nearly half (23). In addition, interspecific microbiomes are more costly to *Nasonia* larval growth and pupation than intraspecific microbiomes (24). Similarly, reciprocal maternal symbiont transplants between two wild, sympatric *Ontophagus* dung beetle species caused developmental delay and elevated mortality in non-native hosts that persisted to the next generation (25). Collectively, phylosymbiotic associations that impact host fitness support the premise that hosts are adapted to their native microbiomes rather than non-native microbiomes, although more studies are needed to confirm these associations and effects in captive and wild host populations."

- Furthermore, even if this pattern were to be consistent or (not inconsistent) with functional symbiosis, it seems that this still falls short of a diagnosis of functional symbiosis (which is what I took the authors to mean when they state they are going to present an evaluation)

As mentioned above, we did not originally intend to use any particular phenotype as a diagnosis *sensu strictu* of function. To avoid confusion and misinterpretation of the term "functional phylosymbiosis", we have now explicitly defined it as "when host and/or microbial phenotypes impact or are impacted by phylosymbiotic associations" (lines 123-124 in revised manuscript)

- [LL 151-153] This was a very disappointing section, where the authors again promise something without adequately delivering it. Authors begin by stating that they will propose an "initial, methodological workflow to statistically evaluate its [phylosymbiosis] strength and significance". Unfortunately, description of this workflow or its associated statistics are not seen. The authors go on to describe a particular experiment and report the secondary conclusions/interpretations from other papers *without* any actual formal description of the tests, statistics, criteria, or any other details of the workflow. I understand that this might have been described in the previous work, but (a) the authors promise a description in *this*

paper; (b) as a review paper that claims to formally introduce the framework, such a description does indeed belong in this paper; (c) without a formal description of the workflow, if they are just going to refer or rehash the results of another paper, without any added insight, then I do not see the point of this paper.

We thank the reviewer for the feedback. We originally wrote a more comprehensive review with technical details of statistical methods, terms, and relevant citations. Unfortunately, we were abruptly required to shorten the manuscript due to journal page limits just before submission, and we resultantly removed these details. We carefully considered comments from all the reviewers and added some of these technical details (see response to comments below) into the manuscript as best as possible given the journal page limits. We apologize in advance for not going further into the details but the length restrictions force a balance between introducing the basics of phylosymbiosis for newcomers (which we think this synthesis will fill a gap in) and providing material for experts in the field.

- [LL 234-235] In a review paper establishing a framework like this, the authors should at the very least present references to these measures.

We note the reviewer's comment and have expanded our description on beta diversity measures with appropriate references (lines 223-234 in revised manuscript):

> "*Microbial beta diversity measures.* Microbial beta diversity, which measures dissimilarities in microbial composition and structure across host samples, is conventionally used to measure phylosymbiosis. Binary measures, such as Jaccard distance and Sørensen-Dice distance (45,46), are calculated with OTU presence/absence data. Quantitative descriptors of OTU abundances can also compute beta diversity, including the Bray-Curtis dissimilarity (47) derived from Motyka et al.'s coefficient (48). It simplifies as 1-[2w/(a+b)], in which w is the sum of the minimum abundances of common species across two host samples, a is the sum of the abundance of all OTUs/species in one sample, and b is the sum of the abundance of all OTUs/species in the other. Phylogeny-based metrics, such as weighted and unweighted unique fraction (UniFrac), use phylogenetic distances between communities (samples) to calculate microbial community differences, necessitating the use of a phylogenetic tree as input (49)."

- [LL 238-239] "are preferred estimators" --- clarify reasoning

In line with reviewer #1's suggestion, we removed our recommendation of quantitative metrics as "preferred estimators". Instead, we have expanded our section on microbial beta diversity measures to include points to be taken into consideration when selecting beta diversity metrics for phylosymbiosis studies (lines 236-247 in revised manuscript):

> "Because beta diversity metrics reflect different aspects of dissimilarity, the choice of metric is study-specific and depends partly on the microbial composition and evolutionary history of the lineages studied. Binary metrics based on presence/absence are more sensitive to variations in rare taxa and were implemented to study host-specificity of sponge microbiomes, where rare taxa comprised more than 90% of distinct OTUs (50). Binary metrics may also be sensitive to recent microbial diversification because recently diverged OTUs/ASVs will exert the same effect as OTUs/ASVs with a longer divergence history (39). In contrast, quantitative metrics are more sensitive to variations in abundant taxa. Besides taxonomy-based phylosymbiosis studies (23,51-53), quantitative metrics have also been applied to metagenomics data (42,43). Metrics that consider phylogenetic relationships between OTUs, such as UniFrac distances, (54) are applied in many other phylosymbiosis studies, including bats (55), corals (20), and mammals (4,43)."

- [LL 276-280] The authors refer to a set of scripts to evaluate the statistical significance of the RF and Matching Cluster distances, and, again, as throughout this paper, defer any explanation, description, and justification of the actual statistics to the frequently-cited Brooks et al. (2017) paper. Again, a reader of this review would not understand much without reading the Brooks et al. 2017 paper, and, conversely, a reader who has read that paper would not gain much at all from reading this paper.

We apologize for the lack of clarity in these statements. Originally, we provided a step-by-step description on how to evaluate statistical significance, but had to suddenly remove this in the final submission due to severe length restrictions by the journal. We added this back to the revised manuscript (lines 285-291) so that readers can understand the process without reading the Brooks et al. 2016 paper:

> "Statistical significance (p-values) is typically evaluated by determining the probability of 100,000 randomized bifurcating dendrogram topologies yielding equivalent or more congruent phylosymbiotic patterns than the microbiome dendrogram (23); normalized Robinson–Foulds and Matching Cluster scores can be calculated as the number of differences between the two topologies divided by the total possible congruency scores for the two trees, with normalized distances ranging from 0 (complete congruence) to 1 (complete incongruence) (23)."

- [LL 276-280] Furthermore, the scripts referenced to do this calculations are poorly documented. The program design is extremely primitive (with, for e.g., paths hard-coded into the scripts and users expected to modify the scripts from run to run for different projects), and there is no indication of any test suites to check for validity of input/output. Furthermore, some of the scripts are simply broken (e.g., one the scripts has the boilerplate copyright text uncommented, resulting in the Python interpreter attempting to execute them). I am aware that the authors themselves are not the primary maintainer of the scripts, nor are the scripts being presented in this paper. However, the authors do reference the scripts as the primary tool by which they evaluate the degree of phylosymbiosis and recommend these scripts to the user.

We apologize for this and removed the reference to the script in the revised manuscript.

- [LL 288-289] "comparing results with those produced by matrix correlation methods". The authors should provide some discussion on the reasons that these methods differ in their approaches and assumptions, and, in particular, what might be captured by one method but not the other. We should have a sophisticated enough appreciation of our statistical approaches not just to throw a million analysis at a problem and hope that they all converge. We pick our methods carefully to address the questions and problems at hand, and when using multiple methods with pick them so that they complement each other's strengths. What do matrix correlation methods reveal that might be missed by topological congruence methods? What

Thank you. We completely agree with the reviewer and removed "comparing results with those produced by matrix correlation methods" in the revised manuscript. We have responded to the reviewer's second point regarding details of matrix correlation methods in our next comment (see below).

- [LL 291-298] This description of matrix correlation methods is disappointingly inadequate for the declared scope of this review paper. A laundry list of methods are mentioned by name as having "been implemented in phylosymbiosis studes" with, again, absolutely no description, characterization, interpretation, or any insight about these methods at all.

Again, we apologize for the omission of details, which we had to suddenly remove due to length restrictions by the journal. Keeping the word limit in mind, we added brief descriptions and advantages of matrix correlation methods to the revised manuscript (lines 304-323):

"Although both topology-based and matrix-based tests are specific and sensitive enough to detect phylosymbiosis in a variety of empirical cases, there are several differences between them. Topological comparison metrics do not use branch length information as there is no a priori reason to assume rates of host evolution in each lineage should equal rates of ecological community change in the microbiome. Indeed, rates of microbiome change may be expected to be far more rapid than gradual evolution of host genetic changes. As such, tests of topology without relative branch lengths are conservative relative to matrix correlation methods that directly rely on comparisons of host genetic divergence with microbial community dissimilarity. A simulation analysis suggested that the Mantel test has higher sensitivity and power than the Robinson-Foulds metric when phylosymbiosis is based on the assumption of microbial preferences for a host trait (19). The practical relevance of this conclusion is not clear because phylosymbiosis will arise from reasons other than microbial colonization preferences, such as host preferences, neutral processes, and microbe-microbe interactions. Moreover, the performance between Mantel test and the more sensitive topology-based Matching Cluster distance was not evaluated in this simulation, and such comparisons are likely to yield different insights. Systematic benchmarking of type I and II error rates of phylosymbiosis measurement methods across various possible scenarios will aid experimental design and result interpretation. As such, research opportunities for the development and implementation of improved phylosymbiosis detection methods are ample."

Referee: 3

Comments to the Author(s)
This review is very well written, it summarizes the main evidence supporting the phylosymbioisis concept across diverse species and with specific examples on how phylosymbioisis shapes host-microbiome associations and fitness across living systems. It also summarizes methodological steps to characterize phylosymbiosis.

We thank the reviewer for the positive feedback.

I only have 2 suggestions on additional minor additions (sections) that would make a more complete review.

1. How are current meta-OMIC techniques to characterize the microbiome and that denote functional rather than taxonomic modules (metagenomics- metabolomics) fit in the phylosymbioisis concept? Most analyses of these type have been based on taxonomic microbiome surveys. Also, would we expect microbiome functionality (genes/pathways/metabolites) to be less constrained by host phylogeny than microbiome taxa/phylogeny?

We have added a few lines in the revised manuscript (lines 219-221) to address the reviewer's questions:

"Metagenomic sequence data were used to demonstrate viral phylosymbiosis in *Nasonia* (42) as well as the varying effects of host phylogeny and ecology on the composition and functions of non-human primate gut microbiomes (43,44)."

There is no clear answer on whether microbiome functionality is more or less constrained by host phylogeny than microbiome taxa/phylogeny, and this has not been specifically measured. We think that the relative effects of host phylogeny on microbial composition and functions likely differ across systems or holobionts, due to various evolutionary and ecological processes shaping host-microbiome specificity. For example, host and/or environmental selection of microbial functions and microbial functional redundancy in some holobionts (possibly gut) may impose higher constraints on microbial functions and less constraints on microbiome taxa by host phylogeny. Other holobionts, such as *Hydra* (mentioned in lines 102-105 of the revised manuscript), produce specific armenin antimicrobial peptide, presumably to select for specific

bacterial taxa, so we would probably expect higher constraints on microbial taxa and lower constraints on microbial functions by host phylogeny.

2. Along these lines, a section with specific examples in which phylosymbiosis has not been supported or in which host phylogenetic patterns does not support congruent microbiomes could be important. Specifically, the importance of shared dietary niche in different (less related) hosts, in shaping or rejecting phylosymbiosis could be discussed.

We agree with the reviewer and did include specific examples which phylosymbiosis has not been supported in the "Prevalence of Phylosymbiosis" section of the original submission, where the examples were organized by host groups and habitats (e.g. mammals, insects, marine animals, plants), e.g.:

> "In *Drosophila* flies, phylosymbiosis patterns are either weakly supported (23) or not detected (80) in both lab strains and wild populations." (lines 372-374 in revised manuscript)
>
> "Studies focusing on gut microbiomes of specific animal groups detected phylosymbiosis in American pikas (51) and *Peromyscus* deer mice (23,52), no phylosymbiosis in western chipmunks (82), and mixed evidence of phylosymbiosis in primates (17,43,44,70), bats (55,83), and birds (62,68,84,85). Besides gut or fecal microbiomes, animal surface microbiomes have also been analyzed for phylosymbiotic associations (86), which for example occur on mammalian skin (53) and passerine feathers (71), but not on amphibian skin (3). (lines 382-388 in revised manuscript)
>
> "Contrarily, qualitative incongruency between Brassicaceae host phylogeny and their root microbiomes has been observed (97), whereas non-statistically significant phylosymbiotic correlations have been reported in other plant microbiome studies (16,98)." (lines 417-420 in revised manuscript)

Appendix B

VANDERBILT UNIVERSITY

Seth R. Bordenstein
Centennial Endowed Chair in Biological Sciences
Departments of Biological Sciences and
Pathology, Microbiology, and Immunology
Associate Director of Vanderbilt Institute for
Infection, Immunology, and Inflammation
Director of the Vanderbilt Microbiome Initiative

January 31, 2020

Dear Dr. Hans Heesterbeek and Colleagues,

Thank you for the positive feedback on our revised manuscript, RSPB-2019-2900. We made minor revisions to our manuscript to address a few remaining issues raised by the reviewer. We hope the revised version is acceptable for publication.

Seth R. Bordenstein
Vanderbilt University

Phone: 615.322.9087
Email: s.bordenstein@vanderbilt.edu

Comments to Author:

Associate Editor Board Member

Thank you for your revised contribution to the special issue. The manuscript has been reviewed by myself and one of the original reviewers, and we both find the work both greatly improved and suitable for publication pending minor revisions. We think the paper will make a nice contribution to the field as a primer for those entering this fast moving topic.

We thank the associate editor for the encouraging feedback and made minor revisions to our manuscript to address the reviewer's comments below. In the revised manuscript, we also cited and briefly mentioned a recent large-scale phylosymbiosis study on animals (lines 388-390).

Referee: 1

The revised version of this review is substantially improved. I think the authors did a good job adjusting the language to be more in line with an achievable scope. It now reads to me as a useful starting point for researchers intersted in the associated methods to identify relevant literature, rather than seeking to be a comprehensive set of recommendations. I have a few small remaining quibbles and suggestions.

We thank the reviewer for the positive feedback and are happy that the manuscript now serves its intended purpose of a starting point for researchers interested in phylosymbiosis.

109-110: On first impression, this example might not seem like phylosymbiosis: here, experimental manipulation of a community induces genomic changes in separate experimental groups of host species. To someone unfamiliar with phylosymbiosis, this could seem backwards. If you have the space, explicitly calling out the fact that the pattern or correlation could arise this way, too, might help solidify the concept for the reader.

We agree with the reviewer and added a statement in lines 110-113 of the revised manuscript to explicitly connect the findings of the *Drosophila* study to the phylosymbiosis concept:

> "This suggests that rather than being passive agents of phylosymbiosis, microbial communities have the potential to induce host genomic changes that could in turn impact the establishment, maintenance, or breakdown of phylosymbiosis."

120: 'evolution-informed' or 'evolutionarily informed' rather than evolutionary-informed

Changed to "evolutionarily-informed" in line 123 of the revised manuscript.

120-121: I don't think you can really have 'evolutionary selective pressures for phylosymbiosis' -- you can have selective pressures that result in phylosymbiosis.

We note the reviewer's comment and changed the phrase to "evolutionary selective pressures that result in phylosymbiosis" (lines 123-124 in the revised manuscript)

161-163: Rather than cast aspersions on 'early misconceptions' and 'strictly narrow presumptions', I'd suggest just straightforwardly stating that phylosymbiosis does not require or imply these things. This will likely be more clear to a reader unfamiliar with the existing literature.

We agree with the reviewer that a more straightforward statement will benefit the readers. We revised the statement to:

> "Phylosymbiosis does not necessarily imply vertical transmission, mutualistic interactions, or evolutionary splitting from a common ancestor via co-evolution, co-speciation, co-diversification, or co-cladogenesis." (lines 163-165 in revised manuscript).

It might also be worth eliminating the first line, which accounts for almost 15% of this entire section now.

We think the first line is a necessary topic sentence to inform readers what the paragraph is about. However, we also agree that the "What is not phylosmbiosis" section is disproportionately short and have now combined this section with the "What is phylosymbiosis" section into "What is and what is not phylosymbiosis" (lines 78-171 revised manuscript).

228-231: If word count limitation is indeed such a problem for this review, is restating the formula for the Bray-Curtis metric really worthwhile?

We included the formula for the Bray-Curtis metric in our previous submission to address reviewer #2's concern on the lack of technical details. However, we agree that with the word count limitation, these lines are unnecessary. They are now removed from the manuscript.

232-234: UniFrac is really using the phylogenetic distances between microbes to estimate a phylogenetic distance between communities. It might also be worth specifying in 234 that the use of a phylogenetic tree *of microbes* is necessitated.

We note the reviewer's comment and have now specified the use of a "microbial phylogenetic tree" (highlighted in bold below) in lines 230-234 in the revised manuscript:

> "Phylogeny-based metrics, such as weighted and unweighted unique fraction (UniFrac), use phylogenetic distances between communities (samples) to calculate microbial community differences, necessitating the use of a **microbial phylogenetic tree** as input to calculate the total community distance (49)."

249-250: it seems like an explicit definition is in order here. E.g. "Microbiome distinguishability, or the characteristic of being able to statistically significantly differentiate the communities of host lineages under evaluation, is a prerequisite…"

Is this something that should be done by researchers initially before subsequently performing tests of phylosymbiosis?

We thank the reviewer for the feedback and modified the statement to more explicitly define "microbiome distinguishability" and to recommend microbiome distinguishability testing before phylosymbiosis tests (emphasis in bold below):

> "Microbiome distinguishability, **or the characteristic of being able to significantly differentiate microbial communities of host lineages under evaluation**, is a prerequisite for phylosymbiosis **and should be tested before evaluating the phylosymbiosis prediction that more similar host species harbor more similar microbiomes** (20,23,51-53)." (lines 249-252 in revised manuscript)

275: please note here which tree (host or microbe) is used here.

We clarified the text to "a microbial phylogenetic tree" (lines 277-278 in revised manuscript)

285: If use of both metrics is 'strongly recommended' it seems reasonable to include some sort of justification.

Since we already presented the differences between the Robinson-Foulds and Matching Cluster metrics in the lines preceding this statement (lines 281-286 of revised manuscript), we decided to remove this line and leave the choice of metric to the readers.

287: Again, because the possibility of affecting false positive rate by including conspecifics isn't specifically mentioned earlier, I think it is important to include here at least some mention of how randomization is done. Simply randomizing tip labels without maintaining conspecific relationships can result in a 'significant' result even in the absence of phylosymbiosis. This is such an important and common stumbling block that I think it is important to address directly, rather than through the oblique comment about 'more study being necessary' in the earlier section.

This is a valid point. We clarified the text and added the reviewer's suggestion (both highlighted in bold below) to lines 286-291 in the revised manuscript.

> "Statistical significance (p-values) **has been** evaluated by determining the probability of 100,000 randomized bifurcating dendrogram topologies yielding equivalent or more congruent phylosymbiotic patterns than the microbiome dendrogram (23). **Moving forward, improved randomization techniques that preserve conspecific relationships will be useful in reducing false positives."**

363-364: Wouldn't at least some of these by definition no longer be phylosymbiosis? Aren't linkages between microbial diversity, function, and environmental factors... just plain old microbial ecology?

Our original intention was to show that these microbial ecology methods can be useful in understanding the ecological (among many other) basis of phylosymbiosis. To clarify our intention, we modified this statement to:

> "Because phylosymbiosis may arise from ecological (among many other) forces, these methods can be useful in understanding the various ecological interactions that possibly underlie phylosymbiosis," (lines 359-361 in revised manuscript)

398-399: I'm not sure what is meant by "when host phylogeny is examined given host identity."

We apologize for the confusion and simplified the text accordingly:

> "Two previous studies in sponges showed significant correlations between host phylogeny and phylogenetic signal on microbial beta diversity. (66,67)." (lines 401-403 in revised manuscript)

437: I don't agree with the shorthand here connecting 'universality' and falsifiability. The universality of a pattern has nothing to do with its being testable. It could be the case that every single microbiome on the planet was correlated with host phylogeny, or none of them—the pattern would still be testable.

We thank the reviewer for the feedback and removed the reference to "universality" in the statement:

> "It provides a testable hypothesis, reflects the variation likely to be seen in nature, and is amenable to explanation by mechanisms that require further investigation." (lines 442-443 in revised manuscript)

Rather, the fact that we see differences in the strength and significance of phylosymbiosis across different systems implies that there are underlying differences in the strength or type of mechanisms generating that signal, which could in turn help us design studies that usefully elucidate those differences. The non-universality of the pattern is thus important, not for its testability, but for its utility.

Indeed, and the next statement in this section (now lines 262-264 in revised manuscript) emphasizes the utility of the presence and absence of phylosymbiosis:

> "The determination of whether phylosymbiosis is present or not is a first step preceding further investigations into mechanistic details, such as the nature of species-species associations and the type(s) of ecological and evolutionary genetic processes underpinning phylosymbiosis."

442-444: I don't know that this comparison is useful, especially without any attempt to define dysbiosis, explore how it's been used, and specifically detail its deficiencies as a term vis a vis phylosymbiosis.

We agree that without explicit comparisons between dysbiosis and phylosymbiosis in the manuscript, these lines may not be helpful. They are now removed from the manuscript.

451: if phyosymbiosis is a pattern and not a mechanism, it can't accelerate or decelerate -- or do -- anything. It could be associated with accelerated or decelerated speciation.

We note the reviewer's comment and revised the statement to:

> "Is phylosymbiosis associated with the acceleration or deceleration of host speciation?" (lines 454-455 in revised manuscript)