THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

A bird's eye view of phylosymbiosis: weak signatures of phylosymbiosis among all 15 species of cranes

Brian K. Trevelline, Jahree Sosa, Barry K. Hartup and Kevin D. Kohl

Article citation details
Proc. R. Soc. B **287**: 20192988.
http://dx.doi.org/10.1098/rspb.2019.2988

Review timeline

Original submission:	12 July 2019
1st revised submission:	23 December 2019
2nd revised submission:	18 February 2020
Final acceptance:	21 February 2020

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

Review History

RSPB-2019-1650.R0 (Original submission)

Review form: Reviewer 1 (Seth Bordenstein)

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?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
The microbiome field is often challenged to find repeatable patterns across systems, but phylosymbiosis is emerging as one of the few. This is a very strong article by Trevelline et al. that makes two substantial contributions to this bona fide trend of phylosymbiosis in the literature. First, they provide a well controlled and statistically comprehensive analyses of bird phyloysmbiosis. Second, they crucially raise the profile and importance of using microbial density estimates in gut microbiome / phylosymbiosis studies. Well done! Please find below a few recommendations that I think will be constructive to the paper's presentation and analyses. I hope these are helpful to the team.

1. In the abstract, I recommend considering the following literature in light of using the words "never in birds". You might replace this tone/framing with "rare in birds" as a result.

Kropáčková L, Těšický M, Albrecht T, Kubovčiak J, Čížková D, Tomášek O, et al. Codiversification of gastrointestinal microbiota and phylogeny in passerines is not explained by ecological divergence. Mol Ecol 2017;26(19):5292-5304.

Javurkova VG, Kreisinger J, Prochazka P, Pozgayova M, Sevcikova K, Brlik V, et al. Unveiled feather microcosm: feather microbiota of passerine birds is closely associated with host species identity and bacteriocin-producing bacteria. ISME J 2019 May 24.

2. In the abstract, the results are framed relative to mammals on line 39, but insects show the strongest phylosymbiosis. It's the word "especially" that throws me off here since mammals are known to have a lower phylosymbiosis stat than insects. Consider bringing this knowledge forward.

3. Lines 52-53: Its important to be comprehensive and incorporate the literature from lab-controlled studies as well since this is relevant to your study. See https://cdn.vanderbilt.edu/vu-web/lab-wpcontent/sites/89/2018/10/21031205/Phylosymbiosis_review_preprint.pdf for a recent summary of the literature.

4. Line 93: Since phylogenetic trees require reconstruction of gene sequence evolution, use the word dendrograms as a more appropriate alternative

5. Lines 111-115: Clarify number of males and females per crane species and re-analyze the data with attention to gender effects that could possibly increase phylosymbiosis. Also clarify age

range if known and does this need to be taken into account as an additional variable causing noise in the signal?

6. Line 120: Were there any cells gated to eukaryotic cells?

7. Line 153: Curious as to why archaea were removed? Was it a primer bias concern? The reason I ask is perhaps archaea contribute to the phylosysymbiotic effect and by excluding it, it may generate a weaker signal than if you included the whole cellular community?

8. Line 214: Strongly recommend using Matching Cluster as well as it is a more sensitive tool for detecting phylosymbiosis within subclades (described in PLOS Biology 2016)

9. Line 236: Give yourself a break over the p-value of 0.09. You might replace "but not clade" with a "non-significant trend" or something like that.

10. Lines 267-277: Repeat analyses with other beta-diversity metrics and report in the manuscript.

11. Lines 275-277: Important result for the field; really glad you did this analysis.

12. LIne 329: Mammals have a generally lower phylosymbiotic signficance than insects. I would make this clearer here.

13. Citation 21: I think this is published.

14. Figures: Thank you for the excellent data visualization.

Sincerely,
Seth Bordenstein

Review form: Reviewer 2

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?
Acceptable

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

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?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
Trevelline et al. conduct a microbiome cataloging and phylosymbiosis analysis of the 15 species of crane. The analyses are thorough and the general question is of interest. I think the addition of absolute abundance estimates in a microbiome paper is great. Unfortunately, I have several substantive issues with the manuscript. The biggest issue is the authors' interpretation of their results (and a few questions about methods). Also, the fact that the manuscript did not have figure legends made it difficult to interpret some of the results (particularly the barcode looking figure?).

MAJOR ISSUES
The authors claim this is significant (albeit weak) evidence for phylosymbiosis in birds and I just cannot see that in the data. It looks like the "phylosymbiosis" signal is entirely driven by a single node - G. monacha / G. nigricollis (fig. 5?) in the abundance controlled dataset. If you remove either of these species from the analysis, the signal goes away entirely. How is that phylosymbiosis? It seems equally likely - or perhaps more likely - that those two sister species are responding to captivity in the same way than that their bodies are selecting for the same communities because they're closely related. It may even be that they're physically closer together in their enclosure(s), couldn't it? I think that even with the modification of "significant (but weak)" signal is misleading. It looks closer to "little to no signal" to me.

I'm not sure I understand the RF calculations and significance (for the absolute abundance informed metric). The RF distance in Figure 5 is 24 so the normalized RF is not 0.84 but 0.92. Two of the fifteen nodes would have to be shared on the two trees for the nRF to be 0.84. Is this the wrong figure or is there a problem with the calculation? I also have a very hard time believing that a single node of sister species was significant on a tree with 15 taxa - how was the p-value calculated? I am wondering if it was calculated as randomized trees with RF less than the calculated RF or randomized trees with RF less than or equal to the calculated RF?

Why just use Bray-Curtis for the phylosymbiosis calculations? I don't see why you wouldn't also include weighted and unweighted unifrac UPGMAs as well.

More broadly, I wonder how we can call it phylosymbiosis without using wild animals in their native habitats. Bringing wild animals into captivity and feeding them a single diet seems like it would erode and bias the signal more than clarify the relationship between the microbiome and the host. The authors call this a "well-controlled test of phylosymbiosis" - but without knowing whether those two birds show phylosymbiotic microbiomes in the wild, I can't distinguish this test from potentially strong human-influence.

I am glad the authors have included absolute abundance measures in this microbiome study; this seems to be the direction the field is going. However, there was no discussion about copy number

variation in 16S. (I'm thinking of Kembel et al 2012 "Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance" but there are many other papers that talk about this issue.) Proteobacteria and Firmicutes are known to have greater than 1 copy of 16S - so couldn't multiplying flow results by the relative abundance just exaggerate natural copy number differences in the members of the community? I think this needs to be discussed. I also would like discussion about what the large amount of variance could mean for these data? The confidence intervals are largely overlapping for all the species (fig. 1, but again hard to decipher without figure legends) - how confident are you that these numbers represent real variation in the cranes (versus minute to minute variations in digestion, for example)?

Were there any negative controls in this study? Or any measures to detect/remove contaminants?

Lastly the topic of bird biology. There is shockingly little discussion about cranes in this paper. How many of these species migrate? Where in their annual cycle were these samples taken? Which species are farthest from their native ranges? What do these birds eat in the wild? Why are they at the Foundation? Were they born in captivity? Have they ever received antibiotics? Are they housed together or in individual cages? Etc.

Minor issues:
What is "a phylogenetic “hot-spot” (sensu Anselin 1995) of gut microbial richness"?

There is no talk about sample size variation in the dataset. Were there any differences in the 2 vs 4 sampled species?

There is no talk about the ages of the birds. Age can affect the microbiome, including in birds, so I think this would ideally be included as a potential contributing factor.

Decision letter (RSPB-2019-1650.R0)

21-Aug-2019

Dear Dr Trevelline:

I am writing to inform you that your manuscript RSPB-2019-1650 entitled "A bird’s eye view of phylosymbiosis: weak but significant signatures of phylosymbiosis among all 15 species of cranes" 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 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
mailto: proceedingsb@royalsociety.org

Associate Editor
Comments to Author:
This is an interesting manuscript that investigates the phenomenon of phylosymbiosis in cranes in a controlled study using captive birds of all 15 species. Phylosymbiosis is an emerging area of research in microbiome science, and the manuscipt provides some interesting data and discussion that would be of benefit to the field. The inclusion of gut microbiota density data is applauded, with the caveat that this still doesn't take into account variable 16S rRNA gene copy numbers across taxa (as highlighted by reviewer 2), which should be discussed. Reviewer 2 also highlights some concerns regarding the veracity of the conclusions that can be made from the data, and both reviewers suggest additional analyses that should be performed to increase the strength of the conclusions. However, the study could make some valuable recommendations for further developments in this area.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
The microbiome field is often challenged to find repeatable patterns across systems, but phylosymbiosis is emerging as one of the few. This is a very strong article by Trevelline et al. that makes two substantial contributions to this bona fide trend of phylosymbiosis in the literature. First, they provide a well controlled and statistically comprehensive analyses of bird phyloysmbiosis. Second, they crucially raise the profile and importance of using microbial density estimates in gut microbiome / phylosymbiosis studies. Well done! Please find below a few recommendations that I think will be constructive to the paper's presentation and analyses. I hope these are helpful to the team.

1. In the abstract, I recommend considering the following literature in light of using the words "never in birds". You might replace this tone/framing with "rare in birds" as a result.

Kropáčková L, Těšický M, Albrecht T, Kubovčiak J, Čížková D, Tomášek O, et al. Codiversification of gastrointestinal microbiota and phylogeny in passerines is not explained by ecological divergence. Mol Ecol 2017;26(19):5292-5304.

Javurkova VG, Kreisinger J, Prochazka P, Pozgayova M, Sevcikova K, Brlik V, et al. Unveiled feather microcosm: feather microbiota of passerine birds is closely associated with host species identity and bacteriocin-producing bacteria. ISME J 2019 May 24.

2. In the abstract, the results are framed relative to mammals on line 39, but insects show the

strongest phylosymbiosis. It's the word "especially" that throws me off here since mammals are known to have a lower phylosymbiosis stat than insects. Consider bringing this knowledge forward.

3. Lines 52-53: Its important to be comprehensive and incorporate the literature from lab-controlled studies as well since this is relevant to your study. See https://cdn.vanderbilt.edu/vu-web/lab-wpcontent/sites/89/2018/10/21031205/Phylosymbiosis_review_preprint.pdf for a recent summary of the literature.

4. Line 93: Since phylogenetic trees require reconstruction of gene sequence evolution, use the word dendrograms as a more appropriate alternative

5. Lines 111-115: Clarify number of males and females per crane species and re-analyze the data with attention to gender effects that could possibly increase phylosymbiosis. Also clarify age range if known and does this need to be taken into account as an additional variable causing noise in the signal?

6. Line 120: Were there any cells gated to eukaryotic cells?

7. Line 153: Curious as to why archaea were removed? Was it a primer bias concern? The reason I ask is perhaps archaea contribute to the phylosysymbiotic effect and by excluding it, it may generate a weaker signal than if you included the whole cellular community?

8. Line 214: Strongly recommend using Matching Cluster as well as it is a more sensitive tool for detecting phylosymbiosis within subclades (described in PLOS Biology 2016)

9. Line 236: Give yourself a break over the p-value of 0.09. You might replace “but not clade” with a “non-significant trend” or something like that.

10. Lines 267-277: Repeat analyses with other beta-diversity metrics and report in the manuscript.

11. Lines 275-277: Important result for the field; really glad you did this analysis.

12. LIne 329: Mammals have a generally lower phylosymbiotic signficance than insects. I would make this clearer here.

13. Citation 21: I think this is published.

14. Figures: Thank you for the excellent data visualization.

Sincerely,
Seth Bordenstein

Referee: 2

Comments to the Author(s)
Trevelline et al. conduct a microbiome cataloging and phylosymbiosis analysis of the 15 species of crane. The analyses are thorough and the general question is of interest. I think the addition of absolute abundance estimates in a microbiome paper is great. Unfortunately, I have several substantive issues with the manuscript. The biggest issue is the authors' interpretation of their results (and a few questions about methods). Also, the fact that the manuscript did not have figure legends made it difficult to interpret some of the results (particularly the barcode looking figure?).

MAJOR ISSUES
The authors claim this is significant (albeit weak) evidence for phylosymbiosis in birds and I just cannot see that in the data. It looks like the "phylosymbiosis" signal is entirely driven by a single node - G. monacha / G. nigricollis (fig. 5?) in the abundance controlled dataset. If you remove either of these species from the analysis, the signal goes away entirely. How is that phylosymbiosis? It seems equally likely - or perhaps more likely - that those two sister species are responding to captivity in the same way than that their bodies are selecting for the same communities because they're closely related. It may even be that they're physically closer together in their enclosure(s), couldn't it? I think that even with the modification of "significant (but weak)" signal is misleading. It looks closer to "little to no signal" to me.

I'm not sure I understand the RF calculations and significance (for the absolute abundance informed metric). The RF distance in Figure 5 is 24 so the normalized RF is not 0.84 but 0.92. Two of the fifteen nodes would have to be shared on the two trees for the nRF to be 0.84. Is this the wrong figure or is there a problem with the calculation? I also have a very hard time believing that a single node of sister species was significant on a tree with 15 taxa - how was the p-value calculated? I am wondering if it was calculated as randomized trees with RF less than the calculated RF or randomized trees with RF less than or equal to the calculated RF?

Why just use Bray-Curtis for the phylosymbiosis calculations? I don't see why you wouldn't also include weighted and unweighted unifrac UPGMAs as well.

More broadly, I wonder how we can call it phylosymbiosis without using wild animals in their native habitats. Bringing wild animals into captivity and feeding them a single diet seems like it would erode and bias the signal more than clarify the relationship between the microbiome and the host. The authors call this a "well-controlled test of phylosymbiosis" - but without knowing whether those two birds show phylosymbiotic microbiomes in the wild, I can't distinguish this test from potentially strong human-influence.

I am glad the authors have included absolute abundance measures in this microbiome study; this seems to be the direction the field is going. However, there was no discussion about copy number variation in 16S. (I'm thinking of Kembel et al 2012 "Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance" but there are many other papers that talk about this issue.) Proteobacteria and Firmicutes are known to have greater than 1 copy of 16S - so couldn't multiplying flow results by the relative abundance just exaggerate natural copy number differences in the members of the community? I think this needs to be discussed. I also would like discussion about what the large amount of variance could mean for these data? The confidence intervals are largely overlapping for all the species (fig. 1, but again hard to decipher without figure legends) - how confident are you that these numbers represent real variation in the cranes (versus minute to minute variations in digestion, for example)?

Were there any negative controls in this study? Or any measures to detect/remove contaminants?

Lastly the topic of bird biology. There is shockingly little discussion about cranes in this paper. How many of these species migrate? Where in their annual cycle were these samples taken? Which species are farthest from their native ranges? What do these birds eat in the wild? Why are they at the Foundation? Were they born in captivity? Have they ever received antibiotics? Are they housed together or in individual cages? Etc.

Minor issues:
What is "a phylogenetic “hot-spot” (sensu Anselin 1995) of gut microbial richness"?

There is no talk about sample size variation in the dataset. Were there any differences in the 2 vs 4 sampled species?

There is no talk about the ages of the birds. Age can affect the microbiome, including in birds, so I think this would ideally be included as a potential contributing factor.

Author's Response to Decision Letter for (RSPB-2019-1650.R0)

See Appendix A.

RSPB-2019-2988.R0

Review form: Reviewer 3 (Jonathan Klassen)

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?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author

In this interesting manuscript, Trevelline et al. perform a rigorous and carefully controlled experiment to determine if there is any correlation between the phylogeny of cranes and the composition of their microbiomes (i.e., phylosymbiosis). This is a valuable addition to the literature because bird microbiomes are often very diverse between species and individuals, and so using a group of individuals belonging to the same taxonomic group but inhabiting the same environment and being fed the same food removes much of the natural variation that would obscure the sort of evolutionary signal that the authors are looking for. Technically, the experiments are very well done and, as noted by the other reviewers, the use of absolute abundances is a big plus. I also commend the authors for their honesty in reporting the weak signals that they found, and see strong value in this work even though the result might be construed as "negative" in some respects. I have a few minor comments that are largely editorial in nature.

1. In the introduction, it might be worthwhile accentuating the contrast between this study (taxonomically restricted to cranes) and previous work that spans a much larger taxonomic breadth. If phylosymbiosis is weak in birds, we might expect it to be strongest at the lowest phylogenetic levels and obscured at higher levels. This might explain why this signal is uncommon in those broader studies.

2. L. 74 "more likely" – it isn't clear to me why what follows is more likely than what precedes this statement. These both seem like reasonable hypotheses to me.

3. LL. 81-82: I think the authors should be careful in using "birds" to describe their study that is done only in cranes. This study can determine if "some birds" exhibit phylosymbiosis, but not "all birds". The wording here could ambiguously mean either.

4. L. 127 – What is it about the large intestinal microbiome that makes it suitable for investigations of phylosymbiosis? Why would some other site not be equally suitable?

5. L. 138 – A version number is needed for the BD FACSDiva software. The actual gating parameters should also be specified here so that the experiment is reproducible.

6. L. 151 – How were the 4 blank extractions related to the sampling scheme? It would seem to that 4 blanks all at processed the beginning would be less informative than blanks spread throughout the extractions.

7. L. 176 – Add "diversity" after "alpha"

8. L. 178 – A version number is needed for Phyloseq

9. L. 179 and L. 186 – The meaning of "collapsed" here is ambiguous. The use of "mean" later in each line seems to imply that "averaged" would be clearer.

10. L. 180 – A version number and parameters are needed for ClaaTU. Ref [24] should be added here at the first mention of the technique

11. L. 193 – A version number is needed for Phylosignal

12. L. 213 – A version number is needed for LEfSe

13. LL. 280-283 – A large number of comparisons are implied here, were these controlled for multiple testing? I am unfamiliar with how ClaaTU handles this issue.

14. LL. 289-290 – Some sort of test statistic should support the statement that "There were no differentially abundant phyla across crane species"

15. L. 319 “among all 15 species of cranes” – Although I am not quite as critical as Reviewer #2, I agree with their point that we should be careful to distinguish the observation of a statistically significant result obtained from a dataset from a process or pattern that operates evenly across that entire dataset. Changing the wording here to something like “a statistically significant pattern of phylosymbiosis in this dataset containing all 15 species of cranes” would clarify that this. My highly accurate visual analysis by squinting at Figure 5 (please note the sarcasm) agrees that most of the signal of phylogenetic congruence is because of the 5 Grus spp. at the bottom of the figure, which would explain why the Matching Cluster statistics (that accommodate clade-specific signal) were stronger than the Robinson-Foulds (which do not).

16. L. 357, L. 363, and elsewhere – In a study like this where you report weak statistical signals, I recommend being careful using “significant” without further qualification. I think that there is a logical jump between “statistically significant” and “biologically significant”, and I think further research is still needed before I am fully convinced of the latter. This doesn’t lessen the value of this study, but highlights that it’s a start and not the end.

17. Capitalization of journal names needs to be checked and corrected for many references.

18. Ref. 15 has now been published and so the page and volume numbers need updating.

19. Page/article numbers are missing or incomplete for refs. 19, 22, 23, 32, 44, and 47

20. Article numbers are incorrect for refs. 20 and 24

21. Reference 30 has now been published in Nature Biotechnology

Decision letter (RSPB-2019-2988.R0)

07-Feb-2020

Dear Dr Trevelline:

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 reviewer and the Associate Editor are very positive. The reviewer, however, raises a limited number of smaller issues. We would like to invite you to revise your manuscript to address them.

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
Comments to Author:
The authors have done a thorough job of addressing the reviewers previous comments. They have also re-analysed some elements of the data to strengthen the veracity of their findings, and consequently the manuscript is much improved. Overall this is a high-quality piece of work that provides data and recommendations that will move this area of microbiome research forward.

Reviewer(s)' Comments to Author:

Referee: 3

Comments to the Author(s).
In this interesting manuscript, Trevelline et al. perform a rigorous and carefully controlled experiment to determine if there is any correlation between the phylogeny of cranes and the composition of their microbiomes (i.e., phylosymbiosis). This is a valuable addition to the literature because bird microbiomes are often very diverse between species and individuals, and so using a group of individuals belonging to the same taxonomic group but inhabiting the same environment and being fed the same food removes much of the natural variation that would obscure the sort of evolutionary signal that the authors are looking for. Technically, the experiments are very well done and, as noted by the other reviewers, the use of absolute abundances is a big plus. I also commend the authors for their honesty in reporting the weak signals that they found, and see strong value in this work even though the result might be construed as "negative" in some respects. I have a few minor comments that are largely editorial in nature.

1. In the introduction, it might be worthwhile accentuating the contrast between this study (taxonomically restricted to cranes) and previous work that spans a much larger taxonomic breadth. If phylosymbiosis is weak in birds, we might expect it to be strongest at the lowest phylogenetic levels and obscured at higher levels. This might explain why this signal is uncommon in those broader studies.

2. L. 74 "more likely" – it isn't clear to me why what follows is more likely than what precedes this statement. These both seem like reasonable hypotheses to me.

3. LL. 81-82: I think the authors should be careful in using "birds" to describe their study that is done only in cranes. This study can determine if "some birds" exhibit phylosymbiosis, but not "all birds". The wording here could ambiguously mean either.

4. L. 127 – What is it about the large intestinal microbiome that makes it suitable for investigations of phylosymbiosis? Why would some other site not be equally suitable?

5. L. 138 – A version number is needed for the BD FACSDiva software. The actual gating parameters should also be specified here so that the experiment is reproducible.

6. L. 151 – How were the 4 blank extractions related to the sampling scheme? It would seem to that 4 blanks all at processed the beginning would be less informative than blanks spread throughout the extractions.

7. L. 176 – Add "diversity" after "alpha"

8. L. 178 – A version number is needed for Phyloseq

9. L. 179 and L. 186 – The meaning of "collapsed" here is ambiguous. The use of "mean" later in each line seems to imply that "averaged" would be clearer.

10. L. 180 – A version number and parameters are needed for ClaaTU. Ref [24] should be added here at the first mention of the technique

11. L. 193 – A version number is needed for Phylosignal

12. L. 213 – A version number is needed for LEfSe

13. LL. 280-283 – A large number of comparisons are implied here, were these controlled for multiple testing? I am unfamiliar with how ClaaTU handles this issue.

14. LL. 289-290 – Some sort of test statistic should support the statement that "There were no differentially abundant phyla across crane species"

15. L. 319 "among all 15 species of cranes" – Although I am not quite as critical as Reviewer #2, I agree with their point that we should be careful to distinguish the observation of a statistically significant result obtained from a dataset from a process or pattern that operates evenly across that entire dataset. Changing the wording here to something like "a statistically significant pattern of phylosymbiosis in this dataset containing all 15 species of cranes" would clarify that this. My highly accurate visual analysis by squinting at Figure 5 (please note the sarcasm) agrees that most of the signal of phylogenetic congruence is because of the 5 Grus spp. at the bottom of the figure, which would explain why the Matching Cluster statistics (that accommodate clade-specific signal) were stronger than the Robinson-Foulds (which do not).

16. L. 357, L. 363, and elsewhere – In a study like this where you report weak statistical signals, I recommend being careful using "significant" without further qualification. I think that there is a logical jump between "statistically significant" and "biologically significant", and I think further research is still needed before I am fully convinced of the latter. This doesn't lessen the value of this study, but highlights that it's a start and not the end.

17. Capitalization of journal names needs to be checked and corrected for many references.

18. Ref. 15 has now been published and so the page and volume numbers need updating.

19. Page/article numbers are missing or incomplete for refs. 19, 22, 23, 32, 44, and 47

20. Article numbers are incorrect for refs. 20 and 24

21. Reference 30 has now been published in Nature Biotechnology

Author's Response to Decision Letter for (RSPB-2019-2988.R0)

See Appendix B.

Decision letter (RSPB-2019-2988.R1)

21-Feb-2020

Dear Dr Trevelline

I am pleased to inform you that your manuscript entitled "A bird's eye view of phylosymbiosis: weak signatures of phylosymbiosis among all 15 species of cranes" 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:
Board Member
Comments to Author:
The authors have comprehensively addressed all of the comments raised by the reviewers.

Appendix A

RESPONSE TO REFEREES

Comments to Author:

This is an interesting manuscript that investigates the phenomenon of phylosymbiosis in cranes in a controlled study using captive birds of all 15 species. Phylosymbiosis is an emerging area of research in microbiome science, and the manuscipt provides some interesting data and discussion that would be of benefit to the field. The inclusion of gut microbiota density data is applauded, with the caveat that this still doesn't take into account variable 16S rRNA gene copy numbers across taxa (as highlighted by reviewer 2), which should be discussed. Reviewer 2 also highlights some concerns regarding the veracity of the conclusions that can be made from the data, and both reviewers suggest additional analyses that should be performed to increase the strength of the conclusions. However, the study could make some valuable recommendations for further developments in this area.

Thank you for reviewing our manuscript entitled *"A bird's eye view of phylosymbiosis: weak but significant signatures of phylosymbiosis among all 15 species of cranes".*

The reviewer comments were thorough, insightful, and greatly improved the quality of our manuscript. In this revised manuscript and response to reviewers, you'll find we have addressed the reviewers concerns by:

1. Analyzing alpha diversity metrics with respect to host sex (presented in Supplemental Figure 1).
2. Using random effect models to account the potential impact of host sex (presented in Figure 2 and Supplemental Figure 2).
3. Including new tests of phylosymbiosis for the UniFrac metrics using the Matching Cluster method, which revealed a significant signal of phylosymbiosis among female cranes (summarized in Table 1).
4. Expanding background on study design and crane biology in general.
5. Softening of language used in title and main text.
6. Discussion of 16S rRNA gene copy number variation among bacteria.
7. General improvements to clarity and completeness.

Below you will find a detailed summary of our revisions in response to all reviewer comments.

Thank you again for your time and for considering this manuscript for publication in *Proceedings B.*

Sincerely,
Brian K. Trevelline

Reviewer 1:

***The microbiome field is often challenged to find repeatable patterns across systems, but phylosymbiosis is emerging as one of the few. This is a very strong article by Trevelline et al. that makes two substantial contributions to this bona fide trend of phylosymbiosis in the literature. First, they provide a well controlled and statistically comprehensive analyses of bird phyloysmbiosis. Second, they crucially raise the profile and importance of using microbial density estimates in gut microbiome / phylosymbiosis studies. Well done! Please find below a few recommendations that I think will be constructive to the paper's presentation and analyses. I hope these are helpful to the team*.**

Thank you for the positive feedback.

1. In the abstract, I recommend considering the following literature in light of using the words "never in birds". You might replace this tone/framing with "rare in birds" as a result.

Kropáčková L, Těšický M, Albrecht T, Kubovčiak J, Čížková D, Tomášek O, et al. Codiversification of gastrointestinal microbiota and phylogeny in passerines is not explained by ecological divergence. Mol Ecol 2017;26(19):5292-5304.

Javurkova VG, Kreisinger J, Procházka P, Pozgayova M, Sevcikova K, Brlik V, et al. Unveiled feather microcosm: feather microbiota of passerine birds is closely associated with host species identity and bacteriocin-producing bacteria. ISME J 2019 May 24.

Text now reads: *"This phenomenon, termed phylosymbiosis, has been observed in diverse evolutionary lineages, but has been difficult to detect in birds."*

2. In the abstract, the results are framed relative to mammals on line 39, but insects show the strongest phylosymbiosis. It's the word "especially" that throws me off here since mammals are known to have a lower phylosymbiosis stat than insects. Consider bringing this knowledge forward.

Text now reads:

> *"Though weak compared to mammals (and especially insects), these results provide evidence of phylosymbiosis in birds. We discuss the potential differences between birds and mammals, such as transmission routes and host filtering, that may underlie the differences in the strength of phylosymbiosis."*

3. Lines 52-53: Its important to be comprehensive and incorporate the literature from lab-controlled studies as well since this is relevant to your study. See https://cdn.vanderbilt.edu/vu-web/lab-wpcontent/sites/89/2018/10/21031205/Phylosymbiosis_review_preprint.pdf for a recent summary of the literature.

We now cite additional examples of lab-controlled phylosymbiosis studies:

"Importantly, this phenomenon, known as phylosymbiosis, occurs even when host species are maintained in a common laboratory environment [7-9], suggesting that gut microbial communities assemble deterministically in accordance with evolutionarily-divergent host phenotypes independent of extrinsic environmental variables."

4. Line 93: Since phylogenetic trees require reconstruction of gene sequence evolution, use the word dendrograms as a more appropriate alternative

We now refer to microbial phylogenetic trees as dendrograms throughout this revised manuscript.

5. Lines 111-115: Clarify number of males and females per crane species and re-analyze the data with attention to gender effects that could possibly increase phylosymbiosis. Also clarify age range if known and does this need to be taken into account as an additional variable causing noise in the signal?

The number of males and females for each crane species varied as shown in Supplemental Data S1. Two species in our cohort, *L. leucogeranus* and *G. japonensis*, were comprised entirely of females. To assess the potential influence of sex on our ability to detect patterns of phylosymbiosis, we conducted supplemental analyses using only female cranes. We now present an alternate version of Figure 1 that visually illustrates the similarity between male and female gut microbiota (Supplemental Figure 1). This approach revealed significant signatures of phylosymbiosis only after removing male cranes from our analysis, suggesting that using mixed-sex animal cohorts may prevent detection of phylosymbiosis. Further, we now include sex as a random effect in our PERMANOVA models. These findings are now described in the results section and Table 1. The median age of cranes in our cohort was 19 years old and ranged between 1 and 55 years as shown in Supplemental Data S1 and mentioned in the methods:

"Cranes in this population varied in age (ranging from 1 to 55 years old) and sex (Supplemental Data S1)."

6. Line 120: Were there any cells gated to eukaryotic cells?

All fecal samples were passed through a 5-micron syringe filter, so most of the eukaryotic cells were removed prior to flow cytometry. Nevertheless, we used gating parameters that would have excluded any eukaryotic cells. We have clarified this in the text:

"Gating parameters were designed to exclude non-fluorescent, eukaryotic, and clustered cell complexes".

7. Line 153: Curious as to why archaea were removed? Was it a primer bias concern? The reason I ask is perhaps archaea contribute to the phylosysymbiotic effect and by excluding it, it may generate a weaker signal than if you included the whole cellular community?

We agree that Archaea are important members of the gut microbiome that may contribute to patterns of phylosymbiosis. The same can be said for fungi and

viruses. However, we were concerned about the bias of the standard 16S primers (515F/806R) in the amplification of Archaea. In order to get a complete picture of Archaeal communities, we would need to use an Archaea-specific primer set (e.g., ARC344F/ARC806R). This was simply outside the scope of the study, but something we will consider in the future.

8. Line 214: Strongly recommend using Matching Cluster as well as it is a more sensitive tool for detecting phylosymbiosis within subclades (described in PLOS Biology 2016)

Thank you for this suggestion as it improved our ability to detect phylosymbiosis. We now test for patterns of phylosymbiosis using both Robinson-Foulds and Matching Cluster methods for Bray-Curtis, Unweighted Unifrac, and Weighted UniFrac metrics (results shown in Table 1). The differences between these approaches are now summarized in the methods:

> *"We tested for patterns phylosymbiosis by comparing gut microbiota dendrograms with the crane phylogeny via the Robinson-Foulds and Matching Cluster metrics with 100,000 random trees using a previously published Python script [7]. While both of these methods are acceptable for assessing topological congruency between trees, the Matching Cluster method accounts for incongruences between closely related branches, and is therefore considered a more refined approach for detecting phylosymbiosis [7]. Both of these approaches produce a normalized score between 0 (complete congruence) and 1 (complete incongruence). P-values were determined by the probability of 100,000 randomized dendrogram topologies yielding equivalent or more congruent phylosymbiotic patterns than the actual microbiota dendrogram."*

9. Line 236: Give yourself a break over the p-value of 0.09. You might replace "but not clade" with a "non-significant trend" or something like that.

We and now refer to p-values between 0.05 and 0.1 as either marginally significant or a non-significant trend.

10. Lines 267-277: Repeat analyses with other beta-diversity metrics and report in the manuscript.

In addition to Bray-Curtis metrics, we now test for phylosymbiosis using unweighted and weighted UniFrac (results reported in Table 1).

11. Lines 275-277: Important result for the field; really glad you did this analysis.

Thank you for the positive feedback.

12. LIne 329: Mammals have a generally lower phylosymbiotic signficance than insects. I would make this clearer here.

Text now reads: "Overall, the phylosymbiotic pattern of cranes was weak compared to those observed in mammals (and especially insects) [7]. These differences may be due to the unique life history traits exhibited by avian host species [11, 18]."

13. Citation 21: I think this is published.

This citation has been updated.

14. Figures: Thank you for the excellent data visualization.

Thank you for the positive feedback.

Reviewer 2:

Trevelline et al. conduct a microbiome cataloging and phylosymbiosis analysis of the 15 species of crane. The analyses are thorough and the general question is of interest. I think the addition of absolute abundance estimates in a microbiome paper is great. Unfortunately, I have several substantive issues with the manuscript. The biggest issue is the authors' interpretation of their results (and a few questions about methods). Also, the fact that the manuscript did not have figure legends made it difficult to interpret some of the results (particularly the barcode looking figure?).

See response to specific comments below. We're not sure what happened to the figure legends, but they should have been available for review. In any case, you should have access to them in this revised manuscript.

MAJOR ISSUES

The authors claim this is significant (albeit weak) evidence for phylosymbiosis in birds and I just cannot see that in the data. It looks like the "phylosymbiosis" signal is entirely driven by a single node - G. monacha / G. nigricollis (fig. 5?) in the abundance controlled dataset. If you remove either of these species from the analysis, the signal goes away entirely. How is that phylosymbiosis? It seems equally likely - or perhaps more likely - that those two sister species are responding to captivity in the same way than that their bodies are selecting for the same communities because they're closely related. It may even be that they're physically closer together in their enclosure(s), couldn't it? I think that even with the modification of "significant (but weak)" signal is misleading. It looks closer to "little to no signal" to me.

We agree that in terms of visual appearance, there doesn't seem to be much topological congruence between host and microbial dendrograms. This is reflected by the weak score when using the Robinson-Foulds method (nRF = 0.917). However, patterns of phylosymbiosis can be difficult to detect when using large or complex phylogenetic trees. As suggested by Reviewer 1, we now conduct tests of phylosymbiosis using the Matching Cluster method, which is a more sensitive approach that accounts for incongruences between closely related branches. This approach produced much more congruent normalized scores (nMC = 0.518, P = 0.086) than the RF method, which was strengthened once accounting for microbial density corrections (0.506, P = 0.063). Using this approach, removing *G. monacha* and *nigricollis* would have less of an effect on the phylosymbioitc signal because there are more nodes to work with compared to the RF method (85 versus 12, respectively). Unfortunately, there is no good way to illustrate phylosymbiosis

using matching cluster, so Figure 5 remains unchanged with the exception of added statistical outputs.

To strengthen our claim that there is “significant (albeit weak) evidence for phylosymbiosis”, we conducted an analysis using only female cranes. Sex can strongly affect microbial communities, and thus we expected that reanalyzing our data with respect to a single sex would reduce intraspecific variation and boost the signal of phylosymbiosis. As expected, the Matching Cluster approach revealed significant topological congruence between the host tree and microbial community dendrograms from female cranes using Bray-Curtis and corrected weighted UniFrac metrics. In our opinion, the caveat that phylosymbiosis among cranes can only be detected once eliminating sex effects qualifies as “weak”. We believe this is a more upfront and transparent communication of our findings.

Regarding the question of whether or not this can be considered phylosymbiosis if “two sister species are responding to captivity in the same way” – this would also be expected under the predictions of phylosymbiosis and is not mutually exclusive from the alternative that “their bodies are selecting for the same communities because they're closely related”. As for the question of proximity effects – *G. monacha* and *nigricollis* were housed on opposite sides of the facility and were separated by pens containing other species that were part of this study. So, it would seem unlikely that proximity is driving this pattern.

I'm not sure I understand the RF calculations and significance (for the absolute abundance informed metric). The RF distance in Figure 5 is 24 so the normalized RF is not 0.84 but 0.92. Two of the fifteen nodes would have to be shared on the two trees for the nRF to be 0.84. Is this the wrong figure or is there a problem with the calculation? I also have a very hard time believing that a single node of sister species was significant on a tree with 15 taxa - how was the p-value calculated? I am wondering if it was calculated as randomized trees with RF less than the calculated RF or randomized trees with RF less than or equal to the calculated RF?

Thank you for noticing this issue. It turns out that there was a problem with the RF calculation when using a previously published R script, but the tree in Figure 5 was correct. We apologize for this confusion. We have updated the RF stats throughout this revised manuscript. P-values were calculated by comparing the number of random trees that were equally or more congruent than the actual microbial dendrogram. We now state this in the methods:

> *“P-values were determined by the probability of 100,000 randomized dendrogram topologies yielding equivalent or more congruent phylosymbiotic patterns than the actual microbiota dendrogram”.*

Why just use Bray-Curtis for the phylosymbiosis calculations? I don't see why you wouldn't also include weighted and unweighted unifrac UPGMAs as well.

This revised manuscript now includes tests for phylosymbiosis using Bray-Curtis, weighted, and unweighted UniFrac metrics (summarized in Table 1).

More broadly, I wonder how we can call it phylosymbiosis without using wild animals in their native habitats. Bringing wild animals into captivity and feeding them a single diet seems like it would erode and bias the signal more than clarify the relationship between the microbiome and the host. The authors call this a "well-controlled test of phylosymbiosis" - but without knowing whether those two birds show phylosymbiotic microbiomes in the wild, I can't distinguish this test from potentially strong human-influence.

Phylosymbiosis is defined as “microbial community relationships that recapitulate the phylogeny of their host” (Lim and Bordenstein 2019) regardless of whether animals are in the wild or captivity. However, the gut microbiome is strongly influenced by local environmental factors (e.g., food availability, climate), and thus it can be challenging to detect phylosymbiosis in wild populations. In other words, environmental colonization would actually weaken and confound patterns of phylosymbiosis. It is for this reason that Lim and Bordenstein identified the use of computational and laboratory approaches to elucidate the causes and effects of phylosymbiosis as a major goal for phylosymbiosis research. We now mention this in our introduction:

> *“As noted above, previous attempts to detect phylosymbiosis in birds may have been unsuccessful due to individual variation in environmental conditions and diets, stemming from pronounced interspecific differences in habitat and foraging requirements. It is for these reasons that the use of computational or laboratory methods has been identified as a major goal for future phylosymbiosis studies [21]. Therefore, we sought to control for these potentially confounding factors by characterizing the gut microbiota of all 15 species of cranes (family Gruidae) maintained on identical diets in a shared captive environment.”*

We do agree that using some fecal samples that were collected from the wild and others from captive cranes could introduce a number of confounding variables, but this was not the case in this study – all but two cranes were born in captivity (See Supplemental Data S1) and all samples were collected from captive cranes.

I am glad the authors have included absolute abundance measures in this microbiome study; this seems to be the direction the field is going. However, there was no discussion about copy number variation in 16S. (I'm thinking of Kembel et al 2012 "Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance" but there are many other papers that talk about this issue.) Proteobacteria and Firmicutes are known to have greater than 1 copy of 16S - so couldn't multiplying flow results by the relative abundance just exaggerate natural copy number differences in the members of the community? I think this needs to be discussed. I also would like discussion about what the large amount of variance could mean for these data? The confidence intervals are largely overlapping for all the species (fig. 1, but again hard to decipher without figure legends) - how confident are you that these numbers represent real variation in the cranes (versus minute to minute variations in digestion, for example)?

We agree that variation in16S rRNA gene copy number among microbiota is an important limitation of culture-independent inventories. We now bring this to the attention of the reader:

> *"One caveat to these findings is that a given bacterial species can possess anywhere between 1 and 15 copies of the 16S rRNA gene [51], and thus 16S rRNA relative abundance values should be interpreted with caution [52]. While correcting 16S rRNA relative abundance values by absolute microbial abundance cannot address this fundamental limitation, the significant variation in microbial densities across crane species provide sufficient justification for this approach. Therefore, we argue that future studies of phylosymbiosis should incorporate measurements of microbial density and possibly corrections for 16S rRNA gene copy number."*

We now discuss intraspecific variation in microbial densities within the context of a previous study in mammals:

> *"Consistent with previous studies in mammals [51], we observed significant interspecific differences in microbial density across crane species with substantial variation among conspecifics (Figure 1). Analyzing microbial densities separately for males and females greatly reduced intraspecific variation (Supplemental Figure 1), suggesting that host sex may affect microbial carrying capacity."*

Were there any negative controls in this study? Or any measures to detect/remove contaminants?

Yes, we did sequence negative controls, but we did not computationally remove them from the analysis based on the recommendations of Goodrich et al 2014 and Eisenhofer et al 2018 (full citations below). This is more of an issue in studies on low-abundance communities (e.g., placenta) than in studies that compare fecal inventories. Nevertheless, we now include a supplemental data file (Supplemental Data S3) that compares the bacterial taxonomic composition of 4 DNA extraction kit controls with those of actual crane samples. Overall, there was little overlap between kit controls and crane samples, with 18 crane individuals exhibiting no overlap with kit controls. For the crane samples with ASVs found in kit controls, the relative abundance of these ASVs was extremely small with an average of 1.1% of total reads/sample. Thus, it is unlikely that kit contaminants influenced our results. We now point the reader to information on extraction kit contaminants (full citation below):

> *Additionally, we conducted four 'blank' extractions to account for possible microbial DNA contaminants within commercial kits [26].*

Further, we now justify this approach and point the reader to two paper that explain why computationally removing contaminant ASVs is not recommended.

> *"We detected a total of 29 ASVs in four negative controls. Overall, the bacterial communities of these negative controls exhibited little overlap with crane fecal samples (Supplemental Data S3). Based on previous evidence*

suggesting that computationally removing contaminant ASVs may strongly influence our results and interpretation [31, 32], we simply summarized the occurrence of these reads among kit controls and crane fecal samples (Supplemental Data S3).”

26. Salter S.J., Cox M.J., Turek E.M., Calus S.T., Cookson W.O., Moffatt M.F., Turner P., Parkhill J., Loman N.J., Walker A.W. 2014 Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology* 12(1), 87.

31. Eisenhofer R., Minich J.J., Marotz C., Cooper A., Knight R., Weyrich L.S. 2018 Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends in microbiology*.

32. Goodrich J.K., Di Rienzi S.C., Poole A.C., Koren O., Walters W.A., Caporaso J.G., Knight R., Ley R.E. 2014 Conducting a microbiome study. *Cell* 158(2), 250-262.

***Lastly the topic of bird biology. There is shockingly little discussion about cranes in this paper. How many of these species migrate? Where in their annual cycle were these samples taken? Which species are farthest from their native ranges? What do these birds eat in the wild? Why are they at the Foundation? Were they born in captivity? Have they ever received antibiotics? Are they housed together or in individual cages? Etc*.**

We have expanded the background on our study design, crane biology, husbandry, etc.:

“We focused our study on a captive population of all 15 species of cranes (family Gruidae; Supplemental Data S1) at the International Crane Foundation (Baraboo, Wisconsin, USA). Cranes in this population are intended for education, conservation, and captive breeding purposes, but offer a unique opportunity to test for patterns of phylosymbiosis in the absence of confounding environmental factors. With the exception of two individuals, all cranes were born in captivity (Supplemental Data S1). Some cranes in this population have a history of antibiotic treatment, but not within 2 months of fecal sample collection (Supplemental Data S1). Captive cranes were housed individually (or in breeding pairs) in 15x18 meter outdoor pens with chain-link fencing along each side and grass covered soil as a substrate. Each pen was covered by flight netting and included a 4.2 x 4.2 m indoor enclosure. Cranes are naturally omnivorous and received identical diets in the form of pelleted food (Zeigler Crane Breeder or Maintenance Diet, Gardners, PA, USA; Supplemental Data S2) and fresh water were provided ad libitum in buckets within the enclosures. The birds were exposed to ambient temperature and natural photoperiod. None of the cranes had a history of chronic infectious disease, malnutrition, or husbandry-related problems, and each bird was visually normal according to experienced keepers in the weeks preceding sample collection.”

"We collected fecal samples from a total of 44 individuals across 15 crane species over a 1-week period in October 2017 (range of 2-4 individuals per species; Supplemental Data S1). Cranes in this population varied in age (ranging from 1 to 55 years old) and sex (Supplemental Data S1). All samples consisted of a single, fresh-appearing voided fecal mass collected into Whirl-Pak collection bags (Nasco, Fort Atkinson, WI) with a sterile cotton swab or sterilized tongue depressor. Fecal samples were immediately frozen at -80°C. All fecal samples were collected over the same 10-day period (Supplemental Data S1). While fecal samples generally represent the composite of several gut regions, previous work has shown that avian fecal material best represents the gut microbiota of large intestine [24], and thus are suitable for investigations of phylosymbiosis."

Crane migratory behavior is complex and difficult to generalize. Many subspecies and populations are migratory while others are not. We now mention this in the discussion:

"While crane migratory behaviors vary across species and populations, the reduced relative abundance of this bacterial genus in a short-range migrant suggests that the microbiome could contribute to some aspect of migratory physiology, though these functional roles are currently unknown and warrant further investigation."

Minor issues:

What is "a phylogenetic "hot-spot" (sensu Anselin 1995) of gut microbial richness"?

This has been clarified:

"Using this approach, G. americana exhibited higher gut microbial richness than expected at random (Ii = -0.521, P = 0.013; Figure 1A)."

There is no talk about sample size variation in the dataset. Were there any differences in the 2 vs 4 sampled species?

Unfortunately, sample sizes for each species were dictated by the composition of International Crane Foundation's captive population. All crane species were housed under identical captive conditions regardless of number of individuals per species. We now mention sample size variation in the methods:

"We collected fecal samples from a total of 44 total individuals across 15 crane species over a 1-week period in October 2017 (range of 2-4 individuals per species; Supplemental Data S1)."

There is no talk about the ages of the birds. Age can affect the microbiome, including in birds, so I think this would ideally be included as a potential contributing factor.

The median age of cranes in our cohort was 19 years old and ranged between 1 and 55 years as shown in Supplemental Data S1 and mentioned in the methods (also see response to Reviewer 1's comment above). Unfortunately, there is no way to control for the effect of age in tests of phylosymbiosis and is a limitation of

our study design. Studies using wild birds would have the same limitation since the age of birds would be unknown beyond 2-3 years.

Appendix B

RESPONSE TO REFEREES

Comments to Author:

The authors have done a thorough job of addressing the reviewers previous comments. They have also re-analysed some elements of the data to strengthen the veracity of their findings, and consequently the manuscript is much improved. Overall this is a high-quality piece of work that provides data and recommendations that will move this area of microbiome research forward.

Thank you for reviewing our manuscript entitled *"A bird's eye view of phylosymbiosis: weak signatures of phylosymbiosis among all 15 species of cranes"* for publication in the special issue of *Proceedings of the Royal Society B* on the *"Application of ecological and evolutionary theory to microbiome community dynamics across systems."*

In this revised manuscript and response to reviewers, you'll find we have addressed the reviewers concerns by:

1. Accentuating the major differences between our study and previous studies
2. Providing additional details on the methodology used for flow cytometry, extractions, and statistical analyses
3. Updating results and Figure 4 to reflect a false-discovery rate threshold of 0.2
4. Updating literature cited with up-to-date and complete citations
5. Making general improvements to clarity and completeness

Attached you will find a detailed summary of our revisions in response to all reviewer comments.

Thank you again for your time and for considering this manuscript for publication in *Proceedings B*.

Sincerely,
Brian K. Trevelline
--

Referee: 3

In this interesting manuscript, Trevelline et al. perform a rigorous and carefully controlled experiment to determine if there is any correlation between the phylogeny of cranes and the composition of their microbiomes (i.e., phylosymbiosis). This is a valuable addition to the literature because bird microbiomes are often very diverse between species and individuals, and so using a group of individuals belonging to the same taxonomic group but inhabiting the same environment and being fed the same food removes much of the natural variation that would obscure the sort of evolutionary signal that the authors are looking for. Technically, the experiments are very well done and, as noted by the other reviewers, the use of absolute abundances is a big plus. I also commend the

authors for their honesty in reporting the weak signals that they found, and see strong value in this work even though the result might be construed as "negative" in some respects. I have a few minor comments that are largely editorial in nature.

Thank you for the positive feedback.

1. ***In the introduction, it might be worthwhile accentuating the contrast between this study (taxonomically restricted to cranes) and previous work that spans a much larger taxonomic breadth. If phylosymbiosis is weak in birds, we might expect it to be strongest at the lowest phylogenetic levels and obscured at higher levels. This might explain why this signal is uncommon in those broader studies.***

We now accentuate the contrast between this study and previous studies as suggested:

> *"As noted above, previous attempts to detect phylosymbiosis in birds may have been unsuccessful due to pronounced differences in habitat and diet across a taxonomically-broad cohort spanning several avian orders."*

2. ***L. 74 "more likely" – it isn't clear to me why what follows is more likely than what precedes this statement. These both seem like reasonable hypotheses to me.***

This sentence has now been revised:

> *"While this lack of support for avian phylosymbiosis could be a product of avian evolutionary history (e.g., weight-saving digestive adaptations) [17], it may also be a product of their sensitivity to extrinsic environmental factors or species-specific differences in natural feeding strategies across many avian orders [11, 18]."*

3. ***LL. 81-82: I think the authors should be careful in using "birds" to describe their study that is done only in cranes. This study can determine if "some birds" exhibit phylosymbiosis, but not "all birds". The wording here could ambiguously mean either.***

We have edited this text to "some birds" as requested.

4. ***L. 127 – What is it about the large intestinal microbiome that makes it suitable for investigations of phylosymbiosis? Why would some other site not be equally suitable?***

The purpose of this sentence was to demonstrate that feces is a useful non-invasive sample type for studies on the avian microbiome. We see how our wording could be interpreted as claiming that only the large intestine is suitable for tests of phylosymbiosis, which was not the intended meaning. While we could certainly argue that the large intestine is the best choice for studies of phylosymbiosis, there is not much empirical evidence on this topic. In this revised manuscript, we now state:

> *"Because previous research has shown that feces provides the best non-invasive approximation of the avian microbiome [26], we collected fecal samples from a total of 44 individuals across 15 crane species over a 1-*

week period in October 2017 (range of 2-4 individuals per species; Supplemental Data S1).”

5. *L. 138 – A version number is needed for the BD FACSDiva software. The actual gating parameters should also be specified here so that the experiment is reproducible.*

We have edited the text to include the software version and more details on the gating parameters:

“The flow cytometry analysis of the microbial cells present in each fecal sample was performed using a BD Biosciences LSR II Flow Cytometer with FACSDiva software (version 8). We used three gating parameters to exclude non-fluorescent, eukaryotic, and clustered cell complexes: forward-angle light scatter (FSC), linear side scatter (SSC), and log FITC. FITC versus event counting data plots were gated to distinguish SYBR-positive fluorescent events from non-fluorescent events. FSC-area versus SSC-area data plots were gated to distinguish CountBright internal reference beads from all other recorded events. FSC-height versus FSC-width data plots were gated to identify single prokaryotic cell events. These settings allowed the software to count 10,000 SYBR-positive (at emission wavelength of 530 nm) counting events that were distinguishable as single-cell bacteria and internal reference beads. Instrument, gating, and event recording settings were identical for all samples. Raw flow cytometry data was used to estimate microbial density of fecal samples using the number of internal reference bead counts according to the manufacturer’s instructions.”

6. *L. 151 – How were the 4 blank extractions related to the sampling scheme? It would seem to that 4 blanks all at processed the beginning would be less informative than blanks spread throughout the extractions.*

DNA was extracted from crane fecal samples over a period of 2 days using 2 different extraction kits. We extracted 2 blanks from each kit (for a total of 4) alongside the crane fecal samples. So, yes, the 4 blank extractions were spread throughout the extractions, but the extractions only lasted 2 days. We now specify that blank extractions were conducted alongside fecal extractions:

“Additionally, we conducted four ‘blank’ extractions alongside the fecal extractions to account for possible microbial DNA contaminants within commercial kits [28].”

7. *L. 176 – Add “diversity” after “alpha”*

Edited as suggested.

8. *L. 178 – A version number is needed for Phyloseq*

We now specify Phyloseq version 1.22.3.

9. *L. 179 and L. 186 – The meaning of “collapsed” here is ambiguous. The use of “mean” later in each line seems to imply that “averaged” would be clearer.*

We now use “averaged ASV abundances” rather than collapsed as suggested.

10. ***L. 180 – A version number and parameters are needed for ClaaTU. Ref [24] should be added here at the first mention of the technique***

We now specify ClaaTU version 0.1 and provide the citation at first mention.

11. ***L. 193 – A version number is needed for Phylosignal***

We now specify Phyloseq version 1.2.

12. ***L. 213 – A version number is needed for LEfSe***

We now specify LEfSe version 1.0.

13. ***LL. 280-283 – A large number of comparisons are implied here, were these controlled for multiple testing? I am unfamiliar with how ClaaTU handles this issue.***

ClaaTU is an exploratory tool that does not correct p-value for multiple comparisons. We did not make this clear in the methods and understand that some readers may find this problematic. In this revised manuscript, we address this issue by applying a relatively permissive false-discovery rate threshold of ≤ 0.2 to identify microbial taxa that may explain patterns of phylosymbiosis in our dataset. Likewise, we have updated Figure 4 to reflect only those bacterial taxa that met the new FDR threshold. While this changed the total number of bacterial taxa that were considered statistically significant from 387 to 87, the percentage of bacterial taxa that composed these 87 groups were essentially unchanged: Firmicutes 65->70%, Clostridiales (34->33%), and Lactobacillales (18->17%). Therefore, this new FDR threshold has not significantly influenced our main conclusions and still serves the purpose of demonstrating the utility of the approach for exploring patterns of phylosymbiosis. The FDR corrections are now explained in the methods:

> *“We corrected p-values for multiple comparisons in R using a relatively permissive false-discovery rate [43] threshold of P ≤ 0.2 to identify microbial taxa that may explain patterns of phylosymbiosis among cranes in this dataset”.*

14. ***LL. 289-290 – Some sort of test statistic should support the statement that “There were no differentially abundant phyla across crane species”***

We now support this statement with the cut-off for statistical significance using LEfSe (LDA > 2.0, P > 0.05), which is also now mentioned in the methods.

15. ***L. 319 “among all 15 species of cranes” – Although I am not quite as critical as Reviewer #2, I agree with their point that we should be careful to distinguish the observation of a statistically significant result obtained from a dataset from a process or pattern that operates evenly across that entire dataset. Changing the wording here to something like “a statistically significant pattern of phylosymbiosis in this dataset containing all 15 species of cranes” would clarify that this. My highly accurate visual analysis by squinting at Figure 5 (please note the sarcasm) agrees that most of the signal of phylogenetic congruence is because***

of the 5 Grus spp. at the bottom of the figure, which would explain why the Matching Cluster statistics (that accommodate clade-specific signal) were stronger than the Robinson-Foulds (which do not).

We have revised this statement as suggested:

> *"These interspecific differences in gut microbial community structure yielded a statistically significant pattern of phylosymbiosis in this dataset containing all 15 species of cranes, but only once all male cranes were removed from our analysis."*

16. ***L. 357, L. 363, and elsewhere – In a study like this where you report weak statistical signals, I recommend being careful using "significant" without further qualification. I think that there is a logical jump between "statistically significant" and "biologically significant", and I think further research is still needed before I am fully convinced of the latter. This doesn't lessen the value of this study, but highlights that it's a start and not the end.***

Throughout the manuscript, we now qualify the use of "significant" as "statistically significant" where appropriate.

17. ***Capitalization of journal names needs to be checked and corrected for many references.***

Citations updated as requested.

18. ***Ref. 15 has now been published and so the page and volume numbers need updating.***

Citation updated as requested.

19. ***Page/article numbers are missing or incomplete for refs. 19, 22, 23, 32, 44, and 47***

Citations updated as requested.

20. ***Article numbers are incorrect for refs. 20 and 24***

Citations updated as requested.

21. ***Reference 30 has now been published in Nature Biotechnology***

Citation updated as requested.