THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

Host and parasite traits predict cross-species parasite acquisition by introduced mammals

Annakate M. Schatz and Andrew W. Park

Article citation details
Proc. R. Soc. B **288**: 20210341.
http://dx.doi.org/10.1098/rspb.2021.0341

Review timeline

Original submission:	10 November 2020
1st revised submission:	9 February 2021
2nd revised submission:	10 March 2021
Final acceptance:	1 April 2021

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

Review History

RSPB-2020-2816.R0 (Original submission)

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?
Excellent

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

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 study, authors explore a several related research questions about parasites that infect new host species outside of these species' native ranges. Their analysis leverage authors' experience with the Global Mammal Parasite Database, which seems well posed to address macroecological questions such as this one. This research area requires a definition for what parasites are "acquired" in new (non-native environments), which is inherently difficult barring intensive field work. The authors define "parasite acquisition" in a rather clever yet still conservative manner, by systematically finding the differences between parasites observed in hosts from native ranges vs. those in non-native ranges. They do not make assumptions about parasites that are lost from species in non-native ranges (loss or functional replacement of parasites). I thought their definition for parasite acquisition was itself a very useful, well justified, and well explained, and provides a simple yet important precedent for the research community.

Authors incorporated existing biogeographical data from IUCN terrestrial range maps together with available georeferenced data from GMPD records. I found their decisions to lump ecoregions into zoogeographic regions to be well justified, with the added advantage of simplifying the geographical analyses (e.g., avoiding running several SDMs) while avoiding overreach in their conclusions.

I appreciated the authors' discussion about nuances in model outputs. For example, the PD plots for prevalence were quite noisy before leveling out at higher prevalence values, and in line 316 the authors also note that acquired parasites generally tended to occur at high prevalence, and were generally found in host species that are more closely related. I wondered whether authors have thoughts about how this patterm relates to force of infection and transmission modes across the different parasite types? (A general question, not one that needs to be addressed in the paper, necessarily).

Authors exploration of phylogenetic distance and the acquisition of generalist vs. specialist parasites, and their ensuing discussion of these patterns, were both super interesting (lines 341-353). This was a nice add-on to the finding that the phylo-distance needing to be traversed by the parasite was more important across models than was host range.

The paper was well written overall. I particularly appreciate that it was easy to follow what could've easily become quite a hairy methods section. One small issue is that I found myself repeatedly confused by the term "host community", which I think the authors use to mean the community of host species that are infected by a particular parasite in an area (equivalent to host

range, or host breadth?); whereas I kept interpreting this term as referring to the ecological interactions among hosts that coexist in a community. Despite my own confusion, it is probably fine to leave this as is, unless other reviewers also found this to be confusing.

I thought the figures were great, and especially liked the example map. Are there other similar maps available somewhere? (this is not a suggestion to add them; just curious)

A few minor points for clarification and correction:

- The figure legend for Fig. S3 notes the top four variables, but I think there are only PD plots for the top three variables.

- Line 298 of the main text should cite Fig S3a, not b.

- I was looking for how the authors contextualized their work with respect to previous contributions by Davies and Pedersen, which (I believe) explored the influence of geographic range overlap vs. phylogenetic relatedness in explaining patterns of parasite sharing for at least two different mammal groups (in two separate papers). Were these studies sufficiently different that this exclusion is warranted?

Congratulations on a great contribution. There are some very interesting patterns here that will hopefully spur several follow up analyses.

Review form: Reviewer 2

Recommendation
Accept as is

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

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

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 authors have prepared a brief study about the traits that predict parasite acquisition in invasive hosts. This is one of a decade's worth of studies to use the Global Mammal Parasite Database to these kinds of ends, and while I think the interesting questions left to be answered with those data are dwindling, the work is clearly sufficiently new to merit publication. The manuscript itself is quite polished and I don't think needs any major changes before publication; I've in fact struggled to find any changes or leads I would suggest as a reviewer, as it reads like a finished copy. My only three suggestions would be

- the title is far too broad and should say "introduced species" or "non-native range" or something similar, else it reads as less distinct than previous work from this group

- specify R, not RStudio, as the software - RStudio is just an interface

- Be more transparent in the abstract that only ungulate and carnivore species are included. GMPD is famously only limited to three mammal groups but only two make the cut for focal species; this isn't a particularly broad sample at all, and I realize there's enough replication that it probably doesn't limit the analysis, but it should be presented more up-front.

Review form: Reviewer 3

Recommendation
Major revision is needed (please make suggestions 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.
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
Please see my comments attached. (See Appendix A)

Decision letter (RSPB-2020-2816.R0)

28-Jan-2021

Dear Professor Schatz:

I am writing to inform you that your manuscript RSPB-2020-2816 entitled "Host and parasite traits predict cross-species parasite acquisition by mammals" 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.
4) Data - please see our policies on data sharing to ensure that you are complying (https://royalsociety.org/journals/authors/author-guidelines/#data).

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
Board Member: 1
Comments to Author:
Thank you for giving Proc B the opportunity to consider this fascinating paper. The reviewers agree on the importance and novelty of the question addressed by this MS, but a few issues must be remedied before we can consider a revised version. The most important concerns are the following:

1. Model formulation must be clarified (see detailed comments from Reviewer 3). In the methods section, please specify what the unit of analysis was (rows of the data frame), how many models were run, on what response, with what predictors, and with what sample sizes.

2. Consider paring down figures so that you show only significant effects OR adding statistics to the plots to highlight the significant/important results.

3. Reviewer 3 would like to inspect your code, but is having trouble accessing it. Would you please upload data/outputs and code as separate files?

4. In the Discussion, please expand your coverage of data limitations (see concerns from Reviewer 3).

5. Please either reference previous contributions by Davies and Pedersen or explain why these fall outside the remit of this paper.

6. Title should be revised to better reflect the content of the paper, including terms like "introduced species" or "non-native range".

7. Abstract should be revised to emphasize that only ungulate and carnivore species are included in the dataset.

Reviewer(s)' Comments to Author:

Referee: 1
Comments to the Author(s)
In this study, authors explore a several related research questions about parasites that infect new host species outside of these species' native ranges. Their analysis leverage authors' experience with the Global Mammal Parasite Database, which seems well posed to address macroecological questions such as this one. This research area requires a definition for what parasites are "acquired" in new (non-native environments), which is inherently difficult barring intensive field work. The authors define "parasite acquisition" in a rather clever yet still conservative manner, by systematically finding the differences between parasites observed in hosts from native ranges vs. those in non-native ranges. They do not make assumptions about parasites that are lost from species in non-native ranges (loss or functional replacement of parasites). I thought their definition for parasite acquisition was itself a very useful, well justified, and well explained, and provides a simple yet important precedent for the research community.

Authors incorporated existing biogeographical data from IUCN terrestrial range maps together with available georeferenced data from GMPD records. I found their decisions to lump ecoregions into zoogeographic regions to be well justified, with the added advantage of simplifying the geographical analyses (e.g., avoiding running several SDMs) while avoiding overreach in their conclusions.

I appreciated the authors' discussion about nuances in model outputs. For example, the PD plots for prevalence were quite noisy before leveling out at higher prevalence values, and in line 316 the authors also note that acquired parasites generally tended to occur at high prevalence, and were generally found in host species that are more closely related. I wondered whether authors have thoughts about how this patterm relates to force of infection and transmission modes across the different parasite types? (A general question, not one that needs to be addressed in the paper, necessarily).

Authors exploration of phylogenetic distance and the acquisition of generalist vs. specialist parasites, and their ensuing discussion of these patterns, were both super interesting (lines 341-353). This was a nice add-on to the finding that the phylo-distance needing to be traversed by the parasite was more important across models than was host range.

The paper was well written overall. I particularly appreciate that it was easy to follow what could've easily become quite a hairy methods section. One small issue is that I found myself repeatedly confused by the term "host community", which I think the authors use to mean the community of host species that are infected by a particular parasite in an area (equivalent to host range, or host breadth?); whereas I kept interpreting this term as referring to the ecological interactions among hosts that coexist in a community. Despite my own confusion, it is probably fine to leave this as is, unless other reviewers also found this to be confusing.

I thought the figures were great, and especially liked the example map. Are there other similar maps available somewhere? (this is not a suggestion to add them; just curious)

A few minor points for clarification and correction:

- The figure legend for Fig. S3 notes the top four variables, but I think there are only PD plots for the top three variables.

- Line 298 of the main text should cite Fig S3a, not b.

- I was looking for how the authors contextualized their work with respect to previous contributions by Davies and Pedersen, which (I believe) explored the influence of geographic range overlap vs. phylogenetic relatedness in explaining patterns of parasite sharing for at least two different mammal groups (in two separate papers). Were these studies sufficiently different that this exclusion is warranted?

Congratulations on a great contribution. There are some very interesting patterns here that will hopefully spur several follow up analyses.

Referee: 2
Comments to the Author(s)
The authors have prepared a brief study about the traits that predict parasite acquisition in invasive hosts. This is one of a decade's worth of studies to use the Global Mammal Parasite Database to these kinds of ends, and while I think the interesting questions left to be answered with those data are dwindling, the work is clearly sufficiently new to merit publication. The manuscript itself is quite polished and I don't think needs any major changes before publication; I've in fact struggled to find any changes or leads I would suggest as a reviewer, as it reads like a finished copy. My only three suggestions would be

- the title is far too broad and should say "introduced species" or "non-native range" or something similar, else it reads as less distinct than previous work from this group

- specify R, not RStudio, as the software - RStudio is just an interface

- Be more transparent in the abstract that only ungulate and carnivore species are included. GMPD is famously only limited to three mammal groups but only two make the cut for focal species; this isn't a particularly broad sample at all, and I realize there's enough replication that it probably doesn't limit the analysis, but it should be presented more up-front.

Referee: 3
Comments to the Author(s)
Please see my comments attached.

Author's Response to Decision Letter for (RSPB-2020-2816.R0)

See Appendix B.

RSPB-2021-0341.R0

Review form: Reviewer 3

Recommendation
Major revision is needed (please make suggestions 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?
Excellent

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.
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
The authors have done a very good job of clarifying their model formulation, and the messaging of the paper as a whole is much clearer. It's still a really interesting paper and I like many of the conclusions. I still have two main comments, which are closely linked and remain from my previous review, but I believe them more strongly now that I understand the model formulation:
1. Figures 3-4 remain busy and confusing, and I urge the authors to simplify them and to make it clearer what the message is from each.
2. The authors' treatment of different effects of other explanatory variables across transmission modes and parasite groups (i.e., interaction effects) does not have statistical support (I think), and is instead based on post hoc interrogation of the raw data plots, which weakens the paper substantially. I apologise for being a stickler, but these relatively intricate conclusions about variation across pathogen subgroups are not necessary to make this an interesting paper with useful ecological implications, and the nuance that they introduce waters down the paper's main conclusions, and will make fewer people read and understand it. This assertion is supported by the fact that these interaction effects make almost no appearance in the abstract, but take up dozens of lines in the paper itself. Either test these effects by fitting them in the models, or remove any discussion of them. I vote the latter, because I am not convinced that e.g. the acquisition of bacteria is more sensitive to local prevalence than the acquisition of viruses, and I can't think of an a priori reason why that would be the case. In the best case scenario, testing the interactions explicitly can comprise a good hypothesis testing exercise because concluding that an interaction effect is/is not present across pathogen groups can help to distinguish between different explanations for the observed patterns (e.g. biology versus sampling bias).

Inspecting the raw data to draw conclusions about interaction effects is invalid because, for example, 1) an effect may be detectable across all groups but not detectable within groups purely because of sample size, and eyeballing the data will not be informative enough to tell whether this is the case; 2) the raw data do not account for any of the other variables in the model, so they may have drastically weakened ability to pick up an interaction (or in the worst case the interactions could go in entirely different directions); and 3) while patterns of a main effect may appear to differ across groups, this could be purely random variation that is in fact within the expected regions under the distribution given a main effect and no interaction. In a simple linear model, it can take >10 times as much data to conclusively test an interaction effect compared to two main effects. All of this contributes to a notable possibility that the identified differences across subgroups are not real.

Everything between lines 312-356 is questionable as a result of this issue. I notice that the authors did explicitly test the interaction between phylogenetic distance and parasite specificity, and display the results in Figure S4 (although without citing statistical support for its importance, so this could still just be based on the eyeballing the data), while also referring to the "consistent and distinct relationship between parasite acquisition and mean phylogenetic distance across parasite

types and transmission modes". Why did they not test for the interactions of various effects with parasite group and transmission mode? If they did test this and found no significant effect, I'm afraid lines 312-356 need to be removed because they're not supported. If they did test this and found it significant, it needs to be reported more clearly (and again I apologise for needlessly kicking up a fuss). If there was some reason it couldn't be tested, they need to explain why and then further justify why they are inspecting raw data for patterns. The paper needs lines like "including an interaction effect between parasite group and prevalence improved model fit by X and/or explained X% of variation".

With regards to figure 3-4: convention states that the explanatory variable (e.g. mean phylogenetic distance) goes on the X axis, with the response variable (parasite acquisition) on the Y axis. Portraying these figures this way would make them more interpretable rather than using a boxplot with acquisition on the X axis. I would expect to see a partial dependence plot output or logistic regression-style plot or similar. If the authors remove their assertions about interaction effects with parasite group and transmission mode, this removes the necessity to facet both figures by these variables, making them clearer and easier to interpret, and ultimately making this a more easily readable and citable paper.

Finally, while I appreciate the addition of a bit of discussion of sampling biases, the authors really need to be more explicit: "our results could have been produced if, for example, disease researchers examining an invasive species are more likely to sample them preferentially for locally prevalent parasites, or globally widespread pathogens like <insert example species>".

Minor comments:

- 141: each host's not each hosts'
- Figure S3, panel b): a few of the hosts have predictions for areas with no data. Might be worth avoiding displaying those areas and/or putting `scales = "free_x"` in the facet argument?

Decision letter (RSPB-2021-0341.R0)

02-Mar-2021

Dear Professor Schatz:

Your manuscript has now been peer reviewed and the review has 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 has raised some concerns with your manuscript and 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 (https://royalsociety.org/journals/authors/author-guidelines/#data). 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:
Thank you for your substantial efforts to revise the MS in response to the last round of reviewer comments. The MS has now been re-reviewed by Reviewer 3, who remains concerned about a few statistical and plotting issues. Please ensure that the next version of the MS takes these important concerns into account.

Reviewer(s)' Comments to Author:
Referee: 3
Comments to the Author(s).
The authors have done a very good job of clarifying their model formulation, and the messaging of the paper as a whole is much clearer. It's still a really interesting paper and I like many of the conclusions. I still have two main comments, which are closely linked and remain from my previous review, but I believe them more strongly now that I understand the model formulation:
1. Figures 3-4 remain busy and confusing, and I urge the authors to simplify them and to make it clearer what the message is from each.
2. The authors' treatment of different effects of other explanatory variables across transmission modes and parasite groups (i.e., interaction effects) does not have statistical support (I think), and is instead based on post hoc interrogation of the raw data plots, which weakens the paper substantially. I apologise for being a stickler, but these relatively intricate conclusions about variation across pathogen subgroups are not necessary to make this an interesting paper with useful ecological implications, and the nuance that they introduce waters down the paper's main conclusions, and will make fewer people read and understand it. This assertion is supported by the fact that these interaction effects make almost no appearance in the abstract, but take up dozens of lines in the paper itself. Either test these effects by fitting them in the models, or remove any discussion of them. I vote the latter, because I am not convinced that e.g. the acquisition of bacteria is more sensitive to local prevalence than the acquisition of viruses, and I can't think of an a priori reason why that would be the case. In the best case scenario, testing the interactions explicitly can comprise a good hypothesis testing exercise because concluding that an interaction effect is/is not present across pathogen groups can help to distinguish between different explanations for the observed patterns (e.g. biology versus sampling bias).

Inspecting the raw data to draw conclusions about interaction effects is invalid because, for example, 1) an effect may be detectable across all groups but not detectable within groups purely because of sample size, and eyeballing the data will not be informative enough to tell whether this is the case; 2) the raw data do not account for any of the other variables in the model, so they may have drastically weakened ability to pick up an interaction (or in the worst case the interactions could go in entirely different directions); and 3) while patterns of a main effect may appear to differ across groups, this could be purely random variation that is in fact within the

expected regions under the distribution given a main effect and no interaction. In a simple linear model, it can take >10 times as much data to conclusively test an interaction effect compared to two main effects. All of this contributes to a notable possibility that the identified differences across subgroups are not real.

Everything between lines 312-356 is questionable as a result of this issue. I notice that the authors did explicitly test the interaction between phylogenetic distance and parasite specificity, and display the results in Figure S4 (although without citing statistical support for its importance, so this could still just be based on the eyeballing the data), while also referring to the "consistent and distinct relationship between parasite acquisition and mean phylogenetic distance across parasite types and transmission modes". Why did they not test for the interactions of various effects with parasite group and transmission mode? If they did test this and found no significant effect, I'm afraid lines 312-356 need to be removed because they're not supported. If they did test this and found it significant, it needs to be reported more clearly (and again I apologise for needlessly kicking up a fuss). If there was some reason it couldn't be tested, they need to explain why and then further justify why they are inspecting raw data for patterns. The paper needs lines like "including an interaction effect between parasite group and prevalence improved model fit by X and/or explained X% of variation".

With regards to figure 3-4: convention states that the explanatory variable (e.g. mean phylogenetic distance) goes on the X axis, with the response variable (parasite acquisition) on the Y axis. Portraying these figures this way would make them more interpretable rather than using a boxplot with acquisition on the X axis. I would expect to see a partial dependence plot output or logistic regression-style plot or similar. If the authors remove their assertions about interaction effects with parasite group and transmission mode, this removes the necessity to facet both figures by these variables, making them clearer and easier to interpret, and ultimately making this a more easily readable and citable paper.

Finally, while I appreciate the addition of a bit of discussion of sampling biases, the authors really need to be more explicit: "our results could have been produced if, for example, disease researchers examining an invasive species are more likely to sample them preferentially for locally prevalent parasites, or globally widespread pathogens like ".

Minor comments:
• 141: each host's not each hosts'
• Figure S3, panel b): a few of the hosts have predictions for areas with no data. Might be worth avoiding displaying those areas and/or putting `scales = "free_x"` in the facet argument?

Author's Response to Decision Letter for (RSPB-2021-0341.R0)

See Appendix C.

RSPB-2021-0341.R1 (Revision)

Review form: Reviewer 1

Recommendation
Accept as is

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?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
The authors have done a good job justifying and explaining their results, and I have no further comments. I look forward to seeing the paper published.

Decision letter (RSPB-2021-0341.R1)

01-Apr-2021

Dear Miss Schatz

I am pleased to inform you that your very nice manuscript entitled "Host and parasite traits predict cross-species parasite acquisition by introduced mammals" 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

Data Accessibility section
Please remember to make any data sets live prior to publication, and update any links as needed when you receive a proof to check. It is good practice to also add data sets to your reference list.

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: 1
Comments to Author:
Thank you for doing such a thorough job of addressing reviewer concerns. I am pleased to accept this excellent MS for publication in Proc B.

Board Member: 2
Comments to Author:
(There are no comments.)

Appendix A

Schatz et al present an interesting, novel and well-written study examining the factors driving parasite acquisition in invasive mammal species. Although the study's conclusions are intriguing and potentially important, and I'm excited to have an answer for this question, it's really hard to judge the veracity of the analysis because of the lack of clarity concerning the model formulation. The model makeup is unusual, which makes the analysis interesting but necessitates better description to properly understand it. For example, what exactly was the unit of analysis? Were the models run separately on different parasite taxa, on different hosts? When were interactions fitted, and with which effects? Was each parasite in the novel area a unique data point, with its acquisition 0/1 as the response variable? I think all of this information is in there in some format, but the difficulty discerning it makes it difficult to assess the potential roles of e.g. sampling and pseudoreplication in producing the results. To exacerbate the situation, the figures provide a deluge of information that really obscures the main points of the paper (see other comments). I think readers will be tripped up by these same things, so I think it will benefit the paper substantially to make these things more transparent.

Given that the paper relies entirely on GMPD as a static data source, it's possible that the effect of prevalence on parasite acquisition could be driven entirely by the fact that researchers are more likely to investigate the most prevalent pathogens in their area, and it's very difficult to conceive of a true negative that could be used to weigh against this effect. The same is possibly true of the "number of hosts globally" effect. These could be serious drawbacks to the analysis, but again it is hard to tell without a more explicit description of the model. My main request is therefore that the authors include a prominent section at the very start of the methods detailing: what the unit of analysis was (rows of the data frame), how many models were run, on what response, with what predictors, and with what sample sizes.

Other comments:

- The authors should investigate and explicitly test whether predictors actually vary by parasite rather than fitting separate models for them by default and assessing which effects were significant for which parasites' models. This approach risks interpreting "significant for one" and "not significant for another" as "significantly different effects", which is a fallacy. For example, if P = 0.049 and P = 0.051 for a given effect on two different parasites, the effect is very unlikely to be substantially different across them, but under the separate models approach it might be (over-)interpreted as such. Fitting the parasites together and then adding an interaction with parasite group and investigating its significance or effect on the model is a much more robust way to investigate variation in slopes or effects across parasite taxa, and less prone to false positives. (If I have understood the model formulations correctly,) Doing the analysis this way will also potentially:
    - Inform whether the differences are more likely to be driven by biological differences or sampling effects.
    - Lead to bigger sample sizes, allowing testing of more host species.
- If I've misinterpreted the model formulations as is, I apologise for harping on about this, but it accentuates the need for more transparency in the models' formulations.
- Related, the figures' comprehensive faceting of all the different drivers to different parasite groups, transmission modes, etc. is overwhelming, and further contributes to the difficulty distinguishing what's happening in the models. I recommend paring the figures down and only displaying relevant/significant interactions and/or adding statistics to the plots to denote which

are important and worth looking at. If more room is needed, Figure 1 takes up an unnecessary amount of space to show very little information. A mosaic plot would show this information more easily, and in one panel, as would a Sina plot or violin plot with widths standardised to the number of data points per parasite group.

- In contrast, Figure 2 is great and very easily interpretable.
- Please upload the data/outputs and code separately. I wanted to inspect the code to answer some of my questions, but the zip file is nearly 1GB which is a bit prohibitive.
- There's a notable absence of Carlson et al's preprint on climate change-associated migrations and novel pathogen sharing. I suggest adding it as a citation in the introduction and discussing how the authors' approach here could build on and potentially improve their framework.
- The discussion should include more extensive discussion about data limitations, beyond the paragraph beginning "We acknowledge": What if the study's results are all driven by sampling effort focussed on highly prevalent native pathogens and globally expansive ones? The discussion sentence addressing this currently plays this off by saying that GMPD is "comprehensive in design, with error-checking steps included", but the fact that this database contains a (once-)comprehensive selection of known parasite records doesn't mitigate the risk of sampling bias driven by preferential sampling of prevalent and widespread pathogen species. Conflating **methodological rigour** (which mostly increases breadth and precision) with **bias remedying** (which increases accuracy) is a common conceptual error in studies like this and it should be avoided here. In essence, this paragraph in the discussion as it currently stands appears to have been inserted to allay concerns about the data, but it doesn't do any substantive discussion of **how** these data biases could be driving the observed patterns.
- Should also discuss:
    - Reliance on GMPD alone, whether there are other data sources that could be used (the Cohen et al. Science paper's prevalence dataset, for example?)
    - Use of only 11 hosts, mostly ungulates. This really does reduce the potential generality of the analysis, despite their apparent ability to cross-predict.

Appendix B

Response to Referees
Host and parasite traits predict cross-species parasite acquisition by introduced mammals
Annakate Schatz and Andrew Park

Associate Editor
Board Member: 1
Thank you for giving Proc B the opportunity to consider this fascinating paper. The reviewers agree on the importance and novelty of the question addressed by this MS, but a few issues must be remedied before we can consider a revised version.

Thank you for your feedback and for the opportunity to revise and resubmit our manuscript.

The most important concerns are the following:

1. Model formulation must be clarified (see detailed comments from Reviewer 3). In the methods section, please specify what the unit of analysis was (rows of the data frame), how many models were run, on what response, with what predictors, and with what sample sizes.

Please see our response to Reviewer 3 for details.

2. Consider paring down figures so that you show only significant effects OR adding statistics to the plots to highlight the significant/important results.

Please see our response to Reviewer 3 for details.

3. Reviewer 3 would like to inspect your code, but is having trouble accessing it. Would you please upload data/outputs and code as separate files?

Please see our response to Reviewer 3 for details.

4. In the Discussion, please expand your coverage of data limitations (see concerns from Reviewer 3).

Please see our response to Reviewer 3 for details.

5. Please either reference previous contributions by Davies and Pedersen or explain why these fall outside the remit of this paper.

We added citations and a brief comment on lines 53-54 of the Introduction.

6. Title should be revised to better reflect the content of the paper, including terms like "introduced species" or "non-native range".

We amended the title to specify introduced mammals as our focus: "Host and parasite traits predict cross-species parasite acquisition by introduced mammals."

7. Abstract should be revised to emphasize that only ungulate and carnivore species are included in the dataset.

We added note of this fact on lines 20-21 of the abstract.

Referee: 1

In this study, authors explore a several related research questions about parasites that infect new host species outside of these species' native ranges. Their analysis leverage authors' experience with the Global Mammal Parasite Database, which seems well posed to address macroecological questions such as this one. This research area requires a definition for what parasites are "acquired" in new (non-native environments), which is inherently difficult barring intensive field work. The authors define "parasite acquisition" in a rather clever yet still conservative manner, by systematically finding the differences between parasites observed in hosts from native ranges vs. those in non-native ranges. They do not make assumptions about parasites that are lost from species in non-native ranges (loss or functional replacement of parasites). I thought their definition for parasite acquisition was itself a very useful, well justified, and well explained, and provides a simple yet important precedent for the research community.

Authors incorporated existing biogeographical data from IUCN terrestrial range maps together with available georeferenced data from GMPD records. I found their decisions to lump ecoregions into zoogeographic regions to be well justified, with the added advantage of simplifying the geographical analyses (e.g., avoiding running several SDMs) while avoiding overreach in their conclusions.

I appreciated the authors' discussion about nuances in model outputs. For example, the PD plots for prevalence were quite noisy before leveling out at higher prevalence values, and in line 316 the authors also note that acquired parasites generally tended to occur at high prevalence, and were generally found in host species that are more closely related. I wondered whether authors have thoughts about how this pattern relates to force of infection and transmission modes across the different parasite types? (A general question, not one that needs to be addressed in the paper, necessarily).

Yes, this is a great point. We now conclude the Discussion paragraph on prevalence with exactly the point that these statistical findings could support future work at the mechanistic level (lines 341-344).

Authors exploration of phylogenetic distance and the acquisition of generalist vs. specialist parasites, and their ensuing discussion of these patterns, were both super interesting (lines 341-353). This was a nice add-on to the finding that the phylo-distance needing to be traversed by the parasite was more important across models than was host range.

The paper was well written overall. I particularly appreciate that it was easy to follow what could've easily become quite a hairy methods section. One small issue is that I found myself repeatedly confused by the term "host community", which I think the authors use to mean the community of host species that are infected by a particular parasite in an area (equivalent to host range, or host breadth?); whereas I kept interpreting this term as referring to the ecological interactions among hosts that coexist in a community. Despite my own confusion, it is probably fine to leave this as is, unless other reviewers also found this to be confusing.

Thank you for pointing this out. We have added a clarification sentence in the Methods section on lines 144-146.

I thought the figures were great, and especially liked the example map. Are there other similar maps available somewhere? (this is not a suggestion to add them; just curious)

We have not made similar maps for the other focal hosts, but would be happy to create and provide them to you should you wish to examine them further.

A few minor points for clarification and correction:

- The figure legend for Fig. S3 notes the top four variables, but I think there are only PD plots for the top three variables.

Thank you for catching this error. It is corrected in the resubmitted version of our Electronic Supplementary Material.

- Line 298 of the main text should cite Fig S3a, not b.

Thank you for spotting this. It is corrected on what is now line 311.

- I was looking for how the authors contextualized their work with respect to previous contributions by Davies and Pedersen, which (I believe) explored the influence of geographic range overlap vs. phylogenetic relatedness in explaining patterns of parasite sharing for at least two different mammal groups (in two separate papers). Were these studies sufficiently different that this exclusion is warranted?

Thank you for pointing out this oversight. We added citations and a brief comment on lines 53-54 of the Introduction.

Congratulations on a great contribution. There are some very interesting patterns here that will hopefully spur several follow up analyses.

Thank you!

Referee: 2

The authors have prepared a brief study about the traits that predict parasite acquisition in invasive hosts. This is one of a decade's worth of studies to use the Global Mammal Parasite Database to these kinds of ends, and while I think the interesting questions left to be answered with those data are dwindling, the work is clearly sufficiently new to merit publication. The manuscript itself is quite polished and I don't think needs any major changes before publication; I've in fact struggled to find any changes or leads I would suggest as a reviewer, as it reads like a finished copy.

Thank you!

My only three suggestions would be

- the title is far too broad and should say "introduced species" or "non-native range" or something similar, else it reads as less distinct than previous work from this group

We amended the title to specify introduced mammals as our focus: "Host and parasite traits predict cross-species parasite acquisition by introduced mammals."

- specify R, not RStudio, as the software - RStudio is just an interface

This is corrected on line 151.

- Be more transparent in the abstract that only ungulate and carnivore species are included. GMPD is famously only limited to three mammal groups but only two make the cut for focal species; this isn't a particularly broad sample at all, and I realize there's enough replication that it probably doesn't limit the analysis, but it should be presented more up-front.

We added note of this fact on lines 20-21 of the abstract.

Referee: 3

Schatz et al present an interesting, novel and well-written study examining the factors driving parasite acquisition in invasive mammal species. Although the study's conclusions are intriguing and potentially important, and I'm excited to have an answer for this question, it's really hard to judge the veracity of the analysis because of the lack of clarity concerning the model formulation. The model makeup is unusual, which makes the analysis interesting but necessitates better description to properly understand it. For example, what exactly was the unit of analysis? Were the models run separately on different parasite taxa, on different hosts? When were interactions fitted, and with which effects? Was each parasite in the novel area a unique data point, with its acquisition 0/1 as the response variable? I think all of this information is in there in some format, but the difficulty discerning it makes it difficult to assess the potential roles of e.g. sampling and pseudoreplication in producing the results. To exacerbate the situation, the figures provide a deluge of information that really obscures the main points of the paper (see other comments). I think readers will be tripped up by these same things, so I think it will benefit the paper substantially to make these things more transparent.

Thank you for highlighting these potential points of confusion for readers. We address each of these concerns in more detail below.

Given that the paper relies entirely on GMPD as a static data source, it's possible that the effect of prevalence on parasite acquisition could be driven entirely by the fact that researchers are more likely to investigate the most prevalent pathogens in their area, and it's very difficult to conceive of a true negative that could be used to weigh against this effect. The same is possibly true of the "number of hosts globally" effect. These could be serious drawbacks to the analysis, but again it is hard to tell without a more explicit description of the model. My main request is therefore that the authors include a prominent section at the very start of the methods detailing: what the unit of analysis was (rows of the data frame), how many models were run, on what response, with what predictors, and with what sample sizes.

Thank you for the suggestion to add an overview paragraph to the Methods section. We agree this will help orient the reader. The requested paragraph is on lines 141-150. We further clarified our

sample sizes in the legend for Table 1 (lines 213-216). We also note on lines 392-396 that the effects of prevalence and global host count could stem from sampling artefacts.

Other comments:

- The authors should investigate and explicitly test whether predictors actually vary by parasite rather than fitting separate models for them by default and assessing which effects were significant for which parasites' models. This approach risks interpreting "significant for one" and "not significant for another" as "significantly different effects", which is a fallacy. For example, if P = 0.049 and P = 0.051 for a given effect on two different parasites, the effect is very unlikely to be substantially different across them, but under the separate models approach it might be (over-)interpreted as such. Fitting the parasites together and then adding an interaction with parasite group and investigating its significance or effect on the model is a much more robust way to investigate variation in slopes or effects across parasite taxa, and less prone to false positives. (If I have understood the model formulations correctly,) Doing the analysis this way will also potentially:
    - Inform whether the differences are more likely to be driven by biological differences or sampling effects.
    - Lead to bigger sample sizes, allowing testing of more host species.

  We believe these concerns stem from a misunderstanding of our model structure. Each model was run on a dataset consisting of multiple parasites – specifically, all of the parasites found in one focal host's non-native range. Effectively, models are per host, rather than per parasite. We believe the new overview paragraph at the start of the Methods section helps to clarify this.
- If I've misinterpreted the model formulations as is, I apologise for harping on about this, but it accentuates the need for more transparency in the models' formulations.

  We appreciate the opportunity to clarify our modeling methods in light of your comments.
- Related, the figures' comprehensive faceting of all the different drivers to different parasite groups, transmission modes, etc. is overwhelming, and further contributes to the difficulty distinguishing what's happening in the models. I recommend paring the figures down and only displaying relevant/significant interactions and/or adding statistics to the plots to denote which are important and worth looking at. If more room is needed, Figure 1 takes up an unnecessary amount of space to show very little information. A mosaic plot would show this information more easily, and in one panel, as would a Sina plot or violin plot with widths standardised to the number of data points per parasite group.

  Thank you for the excellent suggestion for Figure 1. We have amended it to a mosaic plot. Regarding Figures 3 and 4, these are intended as general data visualization, showing the underlying data that went into our models, rather than as plots to display any kind of model output. As such, there are no associated metrics of interaction significance to display. We now clarify this point on lines 317-318.
- In contrast, Figure 2 is great and very easily interpretable.
- Please upload the data/outputs and code separately. I wanted to inspect the code to answer some of my questions, but the zip file is nearly 1GB which is a bit prohibitive.

  We have now uploaded a separate code-only zipfile to the manuscript's associated figshare collection. This affords the reader the opportunity to inspect code without downloading large amounts of data. However, to fully reproduce the work, all data would need to be downloaded.
- There's a notable absence of Carlson et al's preprint on climate change-associated migrations and novel pathogen sharing. I suggest adding it as a citation in the introduction and discussing how the authors' approach here could build on and potentially improve their framework.

  Thank you for alerting us to this interesting preprint. We now cite this work on line 39.
- The discussion should include more extensive discussion about data limitations, beyond the paragraph beginning "We acknowledge": What if the study's results are all driven by sampling effort focussed on highly prevalent native pathogens and globally expansive ones? The discussion sentence addressing this currently plays this off by saying that GMPD is "comprehensive in design, with error-checking steps included", but the fact that this database

contains a (once-)comprehensive selection of known parasite records doesn't mitigate the risk of sampling bias driven by preferential sampling of prevalent and widespread pathogen species. Conflating methodological rigour (which mostly increases breadth and precision) with bias remedying (which increases accuracy) is a common conceptual error in studies like this and it should be avoided here. In essence, this paragraph in the discussion as it currently stands appears to have been inserted to allay concerns about the data, but it doesn't do any substantive discussion of how these data biases could be driving the observed patterns.

We expanded the paragraph on lines 388-404 to address these concerns. On line 392, we note the difficulty of distinguishing patterns driven by data biases. The potential influence of prevalent and widespread pathogen species is noted on lines 395-396. We also added a comment on our use of ecological and phylogenetic distances and how missing data might have biased those predictors (lines 396-400).

- Should also discuss:
    - Reliance on GMPD alone, whether there are other data sources that could be used (the Cohen et al. Science paper's prevalence dataset, for example?)

      Co-author Park (also a co-author of GMPD) has combined datasets in other macroecological studies of infectious diseases. There is a delicate judgement involved in doing this: specifically, the advantage of increasing the data quantitatively and the potential disadvantage of decreasing it qualitatively (due to different search rules, screening criteria and how data summaries are reported). Here, we elect the conservative path of using a dataset we know well. We also note that in terms of terrestrial mammals, the bulk of the Cohen et al. data are from GMPD. Inclusion of other taxa studied in Cohen et al. (e.g. birds, amphibians) is likely to introduce new problems of defining native and invasive host ranges that, while of great interest, are beyond the scope of our study.
    - Use of only 11 hosts, mostly ungulates. This really does reduce the potential generality of the analysis, despite their apparent ability to cross-predict.

      We added an acknowledgement of this data limitation on lines 413-414. We contend that our study is among the first to move beyond case studies and identify general rules for parasite acquisition. Finding support for such rules among hosts from distinct orders is a promising beginning. We agree with the reviewer that building on this in future studies would further add to the generality of the findings.

Appendix C

Response to Referees
Host and parasite traits predict cross-species parasite acquisition by introduced mammals
Annakate Schatz and Andrew Park

Associate Editor

Thank you for your substantial efforts to revise the MS in response to the last round of reviewer comments. The MS has now been re-reviewed by Reviewer 3, who remains concerned about a few statistical and plotting issues. Please ensure that the next version of the MS takes these important concerns into account.

Thank you for the opportunity to further revise this manuscript for Proceedings B.

Referee: 3

The authors have done a very good job of clarifying their model formulation, and the messaging of the paper as a whole is much clearer. It's still a really interesting paper and I like many of the conclusions.

Thank you! We are pleased that our revisions effectively addressed these previous concerns.

I still have two main comments, which are closely linked and remain from my previous review, but I believe them more strongly now that I understand the model formulation:

1. Figures 3-4 remain busy and confusing, and I urge the authors to simplify them and to make it clearer what the message is from each.

   We have simplified both Figures 3 and 4. In Figure 3, we removed the transmission mode faceting to focus on variation across parasite types, which is the more heterogeneous parasite grouping in terms of responses. In Figure 4, we reduced the number of panels from four to two; Figure S4 was also adjusted in the same way.
2. The authors' treatment of different effects of other explanatory variables across transmission modes and parasite groups (i.e., interaction effects) does not have statistical support (I think), and is instead based on post hoc interrogation of the raw data plots, which weakens the paper substantially. I apologise for being a stickler, but these relatively intricate conclusions about variation across pathogen subgroups are not necessary to make this an interesting paper with useful ecological implications, and the nuance that they introduce waters down the paper's main conclusions, and will make fewer people read and understand it. This assertion is supported by the fact that these interaction effects make almost no appearance in the abstract, but take up dozens of lines in the paper itself. Either test these effects by fitting them in the models, or remove any discussion of them. I vote the latter, because I am not convinced that e.g. the acquisition of bacteria is more sensitive to local prevalence than the acquisition of viruses, and I can't think of an a priori reason why that would be the case. In the best case scenario, testing the interactions explicitly can comprise a good hypothesis testing exercise because concluding that an interaction effect is/is not present across pathogen groups can help to distinguish between different explanations for the observed patterns (e.g. biology versus sampling bias).

   We first wish to clarify that our models do fit interactions. Boosted regression trees do this in an automated way; we simply provide the data and specify the interaction degree (3, in this case).

   To support our post hoc data visualizations, we now use the Friedman's H statistic [1] to assess the strength of the interactions shown in Figures 3 and S4. This is described on lines 270-274 of the Methods, with results on lines 332 and 377. The method we use, which has been gaining popularity in ecology for many years, does not provide p-values on interaction terms as would be returned in, say, a GLM. However, we believe the many benefits of this method outweigh such a cost. We have revised the language in the Results & Discussion section to avoid the impression of over-interpreting results concerning interactions (lines 331-388). This combined Results & Discussion section has the opportunity to discuss the many interesting results in step with the presentation of findings, but does necessitate judicious language so that discussion is proportionate to findings, and we're happy to adjust based on reviewer feedback.

Inspecting the raw data to draw conclusions about interaction effects is invalid because, for example, 1) an effect may be detectable across all groups but not detectable within groups purely because of sample size, and eyeballing the data will not be informative enough to tell whether this is the case; 2) the raw data do not account for any of the other variables in the model, so they may have drastically weakened ability to pick up an interaction (or in the worst case the interactions could go in entirely different directions); and 3) while patterns of a main effect may appear to differ across groups, this could be purely random variation that is in fact within the expected regions under the distribution given a main effect and no interaction. In a simple linear model, it can take >10 times as much data to conclusively test an interaction effect compared to two main effects. All of this contributes to a notable possibility that the identified differences across subgroups are not real.

The reviewer makes a fair point that since these visualizations pool data across clusters, the broad trends observed in the figures might not be mirrored in individual models. Thus, we now test for interactions to support our choice of data visualizations, and adjust language towards a more proportionate contextualization of results, as noted above.

Everything between lines 312-356 is questionable as a result of this issue. I notice that the authors did explicitly test the interaction between phylogenetic distance and parasite specificity, and display the results in Figure S4 (although without citing statistical support for its importance, so this could still just be based on the eyeballing the data), while also referring to the "consistent and distinct relationship between parasite acquisition and mean phylogenetic distance across parasite types and transmission modes". Why did they not test for the interactions of various effects with parasite group and transmission mode? If they did test this and found no significant effect, I'm afraid lines 312-356 need to be removed because they're not supported. If they did test this and found it significant, it needs to be reported more clearly (and again I apologise for needlessly kicking up a fuss). If there was some reason it couldn't be tested, they need to explain why and then further justify why they are inspecting raw data for patterns. The paper needs lines like "including an interaction effect between parasite group and prevalence improved model fit by X and/or explained X% of variation".

As noted above, we now calculate the strength of two-way interactions shown in Figures 3 and S4. We recognize that we lack sample size to detect a strong three-way interaction (as shown in Figure 4) and have reworded discussion to remove the implication of compelling statistical support for this (lines 363-376).

With regards to figure 3-4: convention states that the explanatory variable (e.g. mean phylogenetic distance) goes on the X axis, with the response variable (parasite acquisition) on the Y axis. Portraying these figures this way would make them more interpretable rather than using a boxplot with acquisition on the X axis. I would expect to see a partial dependence plot output or logistic regression-style plot or similar. If the authors remove their assertions about interaction effects with parasite group and transmission mode, this removes the necessity to facet both figures by these variables, making them clearer and easier to interpret, and ultimately making this a more easily readable and citable paper.

We appreciate this comment. Because we have trialed various versions of figures in research presentations and received supportive comments on figures from another reviewer, we have elected to keep the figure orientation as-is. As we provide detailed PDPs in Figure S3, it would be redundant to use that output in Figures 3 and 4. A logistic regression-style plot is not appropriate because we do not have a regression line to display so it would just show points at 0 and 1; our plots display the distributions of those points to allow better visual comparison.

Finally, while I appreciate the addition of a bit of discussion of sampling biases, the authors really need to be more explicit: "our results could have been produced if, for example, disease researchers examining an invasive species are more likely to sample them preferentially for locally prevalent parasites, or globally widespread pathogens like ".

Thank you for the clarification. On lines 399-401, we have now added the sentence, "Some of our results could have been produced if, for example, researchers studying an invasive species are more likely to sample preferentially for locally prevalent parasites, or non-random, globally widespread pathogens [2]."

Minor comments:

- 141: each host's not each hosts'
  Thank you for catching this typo. It is fixed on line 141.
- Figure S3, panel b): a few of the hosts have predictions for areas with no data. Might be worth avoiding displaying those areas and/or putting `scales = "free_x"` in the facet argument?
  It is necessary to keep the x-axes as they are in order to standardize the histogram bins across plots. However, we have added a sentence to the figure caption to clarify that the predictions do extend beyond the underlying data. It is not unusual for PDPs to be shown this way; in the field of disease ecology, this practice can be seen in, for example, Fig. S8 from Childs et al [3].

References

1. Friedman JH, Popescu BE. 2008 Predictive learning via rule ensembles. *Ann. Appl. Stat.* **2**, 916–954. (doi:10.1214/07-AOAS148)

2. Byers JE, Schmidt JP, Pappalardo P, Haas SE, Stephens PR. 2019 What factors explain the geographical range of mammalian parasites? *Proc. R. Soc. B Biol. Sci.* **286**, 20190673. (doi:10.1098/rspb.2019.0673)

3. Childs ML, Nova N, Colvin J, Mordecai EA. 2019 Mosquito and primate ecology predict human risk of yellow fever virus spillover in Brazil. *Philos. Trans. R. Soc. B Biol. Sci.* **374**, 20180335. (doi:10.1098/rstb.2018.0335)