THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

# The genetics of morphological and behavioural island traits in deer mice

Felix Baier and Hopi E. Hoekstra

**Article citation details**
*Proc. R. Soc. B* **286**: 20191697.
http://dx.doi.org/10.1098/rspb.2019.1697

**Review timeline**

| | |
|---|---|
| Original submission: | 22 January 2019 |
| 1st revised submission: | 13 August 2019 |
| 2nd revised submission: | 30 September 2019 |
| Final acceptance: | 8 October 2019 |

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

# Review History

## RSPB-2019-0131.R0 (Original submission)

### Review form: Reviewer 1

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

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

**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?**
N/A

**Is it clear?**
N/A

**Is it adequate?**
No

**Do you have any ethical concerns with this paper?**
No

**Comments to the Author**
This paper re-evaluates a classic case of 'island syndrome' focusing on evaluating the genetic basis of morphological and behavioural differentiation between island and mainland Peromyscus species in British Columbia. The work is clearly comprehensive and the authors have paid attention to detail teasing apart the relative impact genetic versus environmental responses may play in morphological and behavioural differences.

Major comments:
While there are limitations to the length of the introduction there are some key components that are missing. In particular, the authors have not really teased apart expectations for selection versus stochastic processes for the 'island syndrome' and differentiated those with expectations of plasticity. Given the direction of the discussion and focus, this should be expanded. There is a lack of detail on the potential role of environment in the second paragraph where the authors focus solely on expectations for genetic differences. I would recommend an addition on the discussion of random processes alongside selection additions in paragraph two to address expectations of plasticity. In addition, the discussion lacks much emphasis on the genetic differences observed, which was well-laid out in the introduction, but not addressed to any extent in the discussion. Stating clear hypotheses in the introduction that are addressed in the discussion may address these concerns.

Body weight - This may be my unfamiliarity with mice specimens and their classification, but it's unclear how the authors can remove specimens based on body/tail length if there is differentiation between island/mainland for these traits. Additional details on the justification for exclusion of specimens and associated references is recommended. In addition, if only using adult male mice - does body size plateau at an age that might not make this a trait confounded with age? Additional detail on how age structure was accounted for in the museum specimens would be beneficial. Finally, for the analysis (Line 142) was year used as a covariate in the model? where year of record might impact estimates of body size depending on food availability.

Figure 1A - How might sampling bias impact the results here? What is the significant context

biologically for a >21g to <16g body size. Colonies from the island were established from multiple sites, but mainland only a single site. How might random processes and historical context have impacted the distribution of trait variances between these two regions? Do the authors have an understanding of potential founder events or population bottlenecks that may have impacted differences? It might be complementary to assess the history of the populations evaluating effective population size. Bayesian coalescence approach may complement 1E to infer the impact historical demographic processes may have on population divergence.

Line 164 & Figure 1C/D - The authors state that the populations are 'genetically separable'. Given that PC1 explains 5.7% of the variance and PC2 explains 3.86% variance I'm wondering how different these values are from random expected differences across groups. The amount of variance explained seems quite low relative to the number of genomic-variants used and it would validate the results to suggest that these differences are more than expected by chance, which is not currently convincing. I'm also surprised by the results from 1D given the low percentage explained in the PC-analyses and I'm wondering what the likelihood is that K=3 as based on the PC results those differences do not seem so clear. At the very least a likelihood of K could be considered in supplemental information.

Minor comments:
Abstract - double-check tenses (Line 9-10)
Line 32 - they = traits evolve
Line 35 - during = following
Line 66 - Clarify that MKRF samples are mainland individuals and how many per individuals were sampled per Saturna, Pender and MKRF.
SupplementPage3-Refers to 54 P. maniculatus samples - but the authors describe 28 Saturna, 9 Pender and 28 mainland - details on sample numbers for different genomic analysis need to be clarified
Line80 - How many male and female individuals for Saturna and MKRF were used to establish colonies?
Line 183 - how many island v. mainland individuals were used in this comparison?
Line 193 - Re-organization of this sentence phrasing statement as a test-able hypothesis would clarify the question asked here.
Justification for methods is placed in the results oftentime (ie: why X-ray measurements) and the manuscript could be re-organized to justify approach within the method over results section.

## 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?**
Good

**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.**
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?**
N/A

**Is it adequate?**
Yes

**Do you have any ethical concerns with this paper?**
No

**Comments to the Author**
See attached file.

## Decision letter (RSPB-2019-0131.R0)

06-Mar-2019

Dear Mr Baier:

I am writing to inform you that your manuscript RSPB-2019-0131 entitled "The genetics of morphological and behavioural island traits in deer mice" 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,
Proceedings B
mailto: proceedingsb@royalsociety.org

Associate Editor
Comments to Author:
There is general agreement that this is a nice paper. However, the referees highlight several issues that require attention. Referee 1 points out that quite a large tranche of the reported work repeats previous studies and this tends to be the best-supported evidence. The novel aspects are more or less questioned in terms of the possibility that some of the observed effects could be explained by bias in the behavioural assays. In my opinion this paper could be publishable but would need considerable reshaping. I would like to see all the analyses that replicate prior work separated clearly from the aspects that are novel, possibly by placing them in supplementary files and referring to them "we confirmed that . . ". This would leave more room to address concerns about interpretation, to add more methodological details and to improve clarity. Lots of useful suggestions have been made. I do not share Referee 2's worry about the levels of genetic distinctness since the populations separate very nicely and there are good aqueous reasons for believing gene flow is absent / minimal. In any revised manuscript a good case will need to be made that experimental bias has been properly controlled for the most interesting, novel aspects.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
This paper re-evaluates a classic case of 'island syndrome' focusing on evaluating the genetic basis of morphological and behavioural differentiation between island and mainland Peromyscus species in British Columbia. The work is clearly comprehensive and the authors have paid attention to detail teasing apart the relative impact genetic versus environmental responses may play in morphological and behavioural differences.

Major comments:
While there are limitations to the length of the introduction there are some key components that are missing. In particular, the authors have not really teased apart expectations for selection versus stochastic processes for the 'island syndrome' and differentiated those with expectations of plasticity. Given the direction of the discussion and focus, this should be expanded. There is a

lack of detail on the potential role of environment in the second paragraph where the authors focus solely on expectations for genetic differences. I would recommend an addition on the discussion of random processes alongside selection additions in paragraph two to address expectations of plasticity. In addition, the discussion lacks much emphasis on the genetic differences observed, which was well-laid out in the introduction, but not addressed to any extent in the discussion. Stating clear hypotheses in the introduction that are addressed in the discussion may address these concerns.

Body weight - This may be my unfamiliarity with mice specimens and their classification, but it's unclear how the authors can remove specimens based on body/tail length if there is differentiation between island/mainland for these traits. Additional details on the justification for exclusion of specimens and associated references is recommended. In addition, if only using adult male mice - does body size plateau at an age that might not make this a trait confounded with age? Additional detail on how age structure was accounted for in the museum specimens would be beneficial. Finally, for the analysis (Line 142) was year used as a covariate in the model? where year of record might impact estimates of body size depending on food availability.

Figure 1A - How might sampling bias impact the results here? What is the significant context biologically for a >21g to <16g body size. Colonies from the island were established from multiple sites, but mainland only a single site. How might random processes and historical context have impacted the distribution of trait variances between these two regions? Do the authors have an understanding of potential founder events or population bottlenecks that may have impacted differences? It might be complementary to assess the history of the populations evaluating effective population size. Bayesian coalescence approach may complement 1E to infer the impact historical demographic processes may have on population divergence.

Line 164 & Figure 1C/D - The authors state that the populations are 'genetically separable'. Given that PC1 explains 5.7% of the variance and PC2 explains 3.86% variance I'm wondering how different these values are from random expected differences across groups. The amount of variance explained seems quite low relative to the number of genomic-variants used and it would validate the results to suggest that these differences are more than expected by chance, which is not currently convincing. I'm also surprised by the results from 1D given the low percentage explained in the PC-analyses and I'm wondering what the likelihood is that K=3 as based on the PC results those differences do not seem so clear. At the very least a likelihood of K could be considered in supplemental information.

Minor comments:
Abstract - double-check tenses (Line 9-10)
Line 32 - they = traits evolve
Line 35 - during = following
Line 66 - Clarify that MKRF samples are mainland individuals and how many per individuals were sampled per Saturna, Pender and MKRF.
SupplementPage3-Refers to 54 P. maniculatus samples - but the authors describe 28 Saturna, 9 Pender and 28 mainland - details on sample numbers for different genomic analysis need to be clarified
Line80 - How many male and female individuals for Saturna and MKRF were used to establish colonies?
Line 183 - how many island v. mainland individuals were used in this comparison?
Line 193 - Re-organization of this sentence phrasing statement as a test-able hypothesis would clarify the question asked here.
Justification for methods is placed in the results oftentime (ie: why X-ray measurements) and the manuscript could be re-organized to justify approach within the method over results section.

Referee: 2

Comments to the Author(s)
See attached file.

## Author's Response to Decision Letter for (RSPB-2019-0131.R0)

See Appendix A.

# RSPB-2019-1697.R0

## Review form: Reviewer 2

**Recommendation**
Accept with minor revision (please list in comments)

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

**General interest: Is the paper of sufficient general interest?**
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.**
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**
See attached file for comments.

## Decision letter (RSPB-2019-1697.R0)

13-Sep-2019

Dear Mr Baier:

Your manuscript has now been peer reviewed and the reviews have been assessed by an Associate Editor. The reviewer's comments (not including confidential comments to the Editor) and the comments from the Associate Editor are included at the end of this email for your reference. As you will see, the reviewers and the Associate Editor have raised some issues 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. Datasets should be deposited in an appropriate publicly available repository and details of the associated accession number, link or DOI to the datasets must be included in the Data Accessibility section of the article (https://royalsociety.org/journals/ethics-policies/data-sharing-mining/). Reference(s) to datasets should also be included in the reference list of the article with DOIs (where available).

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

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

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

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

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

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

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

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

Best wishes,
Professor Hans Heesterbeek
mailto: proceedingsb@royalsociety.org

Associate Editor Board Member
Comments to Author:
I find the authors have done a thorough job of addressing issues raised and / or justifying why changes were not made. I rather like the Referee's suggestion for testing for retained differences in aggression and suggest this is done unless there are good reasons for not doing so. Also, while I agree that inter-year variation is unlikely to create trends, I do think inter-year variation is likely present. One solution would be to fit year as a random factor. This might well reveal stronger trends by removing a chunk of the error variance.

Reviewer(s)' Comments to Author:

Referee: 2

Comments to the Author(s).
See attached file for comments.

## Author's Response to Decision Letter for (RSPB-2019-1697.R0)

See Appendix B.

## Decision letter (RSPB-2019-1697.R1)

08-Oct-2019

Dear Mr Baier

I am pleased to inform you that your manuscript entitled "The genetics of morphological and behavioural island traits in deer mice" 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:
I am happy with the way the outstanding points have been addressed

# Appendix A

## Associate Editor

There is general agreement that this is a nice paper. However, the referees highlight several issues that require attention. Referee 2 points out that quite a large tranche of the reported work repeats previous studies and this tends to be the best-supported evidence. The novel aspects are more or less questioned in terms of the possibility that some of the observed effects could be explained by bias in the behavioural assays. In my opinion this paper could be publishable but would need considerable reshaping. I would like to see all the analyses that replicate prior work separated clearly from the aspects that are novel, possibly by placing them in supplementary files and referring to them “we confirmed that..“. This would leave more room to address concerns about interpretation, to add more methodological details and to improve clarity. Lots of useful suggestions have been made. I do not share Referee 1’s worry about the levels of genetic distinctness since the populations separate very nicely and there are good aqueous reasons for believing gene flow is absent / minimal. In any revised manuscript a good case will need to be made that experimental bias has been properly controlled for the most interesting, novel aspects.

>> Thank you for your guidance here. As you will see below, we address, where possible, all of the reviewers' comments.

Here we address your two main comments: While some of the work is indeed consistent with previous studies (e.g. there are island-mainland body size differences in wild populations), we feel it is important to demonstrate and visually present some of the data from the precise species and populations that we are studying in both the field and the lab to allow the reader to make direct and controlled comparisons between wild and lab-raised mice, rather than relying on patterns or data from previous research that either has small sample sizes, took measurements differently, or conducted different assays altogether. Nonetheless, we took this comment to heart, and have worked hard to streamline these sections, moved some figure panels to the Supplement, and use the language you suggested. We also address the comments about the behavioral assays in a detailed response to Reviewer 2 below.

## Referee: 1

Comments to the Author(s)
This paper re-evaluates a classic case of 'island syndrome' focusing on evaluating the genetic basis of morphological and behavioural differentiation between island and mainland Peromyscus species in British Columbia. The work is clearly comprehensive and the authors have paid attention to detail teasing apart the relative impact genetic versus environmental responses may play in morphological and behavioural differences.

>> We thank the Referee for his/her support of our paper.

Major comments:
While there are limitations to the length of the introduction there are some key components that are missing. In particular, the authors have not really teased apart expectations for selection versus stochastic processes for the 'island syndrome' and differentiated those with expectations of plasticity. Given the direction of the discussion and focus, this should be expanded. There is a lack of detail on the potential role of environment in the second paragraph where the authors focus solely on expectations for genetic differences. I would recommend an addition on the discussion of random processes alongside selection additions in paragraph two to address expectations of plasticity.

>> We agree with the Referee that the introduction is brief and that expectations of genetics vs. plasticity are primarily flushed out in the discussion. Therefore, we have edited the first two paragraphs of the introduction to make the language more balanced to both potential genetic and plastic influences on island traits. *Note*: in this paper, we focus primarily on presenting data to disentangle genetic versus plastic effects, rather than selection versus drift.

In addition, the discussion lacks much emphasis on the genetic differences observed, which was well-laid out in the introduction, but not addressed to any extent in the discussion. Stating clear hypotheses in the introduction that are addressed in the discussion may address these concerns.

>> We discuss the genetic differences, namely the finding of island offspring and maternal genotypes driving larger body size, in lines 351-361 of the discussion.

Body weight - This may be my unfamiliarity with mice specimens and their classification, but it's unclear how the authors can remove specimens based on body/tail length if there is differentiation between island/mainland for these traits. Additional details on the justification for exclusion of specimens and associated references is recommended.

>> Our primary concern with filtering the museum records was to exclude (1) juvenile or subadult specimens, and (2) specimens belonging to other *Peromyscus* species (specifically, *P. keeni*). To address concern (1), we removed specimens with body length < 70mm and/or tail length < 60mm. We chose these thresholds because specimens with such measurements in our field-collected sample - which unlike the museum specimens we were able to examine physically - were clearly not adult (based on overall appearance, including coat color, and reproductive status etc.). To address the second concern, we used standard diagnostic thresholds developed by Zheng et al (2003) for tail/body length and tail length to distinguish *P. keeni* - which has a much longer tail - from any *P. maniculatus*. Tail length, although not perfect, is a key diagnostic trait in rodents, used commonly to distinguish among species.

In addition, if only using adult male mice - does body size plateau at an age that might not make this a trait confounded with age? Additional detail on how age structure was accounted for in the museum specimens would be beneficial.

>> We excluded juvenile and subadult specimens in the way described above because we deemed this to be the greatest possible source of noise in the body weight data. Beyond this age classification, it is not possible to distinguish precise ages among adults from museum records. We don't have any reason to suspect we are systematically collecting mice of different ages from island versus mainland populations. Moreover, while it is true that mice grow continuously, we know from our data based on lab colonies that size increase is minimal once sexual maturity is reached.

Finally, for the analysis (Line 142) was year used as a covariate in the model? where year of record might impact estimates of body size depending on food availability.

>> We did not include year as a covariate in these analyses. Because the 504 records were collected across 37 different years, throughout most of the 20th century (1930 - 2014), we reasoned that biased sampling of year-to-year weight fluctuations is unlikely to explain the major island-mainland trends we observe. Moreover, these island-mainland differences remain when we control for food availability in our laboratory colonies (and this, of course, was one of the motivations for establishing laboratory colonies).

Figure 1A - How might sampling bias impact the results here? What is the significant context biologically for a >21g to <16g body size. Colonies from the island were established from multiple sites, but mainland only a single site. How might random processes and historical context have impacted the distribution of trait variances between these two regions? Do the authors have an understanding of potential founder events or population bottlenecks that may have impacted differences? It might be complementary to assess the history of the populations evaluating effective population size. Bayesian coalescence approach may complement 1E to infer the impact historical demographic processes may have on population divergence.

>> First, regarding the scale of the body weight gradient in Fig. 1A, we use 16g and 21g because they are the medians of the weight distribution of mainland and island mice, respectively (Fig. 1B). These thresholds were thus chosen as the minimum and maximum of the color gradient to visually show the overlap of the distributions. *Note*: We have moved Fig. 1B to the Supplement.

We agree that the history of island and mainland populations in the Strait of Georgia likely affected body size variance in this region (Fig. 1A), including migration patterns, colonization trajectories, potential bottlenecks, and associated demographic variables. A comprehensive answer to this problem would require

broader genetic sampling, which is beyond the scope of this current paper. However, we provide some novel insights about these processes (gene flow, divergence time, Fig. 1B-D) for the focal island and mainland populations. From this, it is clear that genetic divergence between the 3 sites on Saturna Island is minimal, and mice likely form one metapopulation on the island.

Line 164 & Figure 1C/D - The authors state that the populations are 'genetically separable'. Given that PC1 explains 5.7% of the variance and PC2 explains 3.86% variance I'm wondering how different these values are from random expected differences across groups. The amount of variance explained seems quite low relative to the number of genomic-variants used and it would validate the results to suggest that these differences are more than expected by chance, which is not currently convincing. I'm also surprised by the results from 1D given the low percentage explained in the PC-analyses and I'm wondering what the likelihood is that K=3 as based on the PC results those differences do not seem so clear. At the very least a likelihood of K could be considered in supplemental information.

>> We respectfully disagree with the Referee on this point. The % variance explained by the first two principal components is quite high for a genetic PCA of population differentiation, which is a testament to the strong genetic divergence of these populations. % variance in genetic PCAs always tends to be rather low (often < 5%) compared to PCAs of other traits, for example morphology, because of (1) the high number of "traits" (=SNPs) considered and (2) because most of the genetic variation segregates within populations, rather than between. Because the populations clearly fall into three clusters in the genetic PCA, we were not concerned about the choice of K=3 for the admixture analysis; indeed the results of the admixture analysis suggest an almost complete absence of gene flow. (Also see remarks from the Associate Editor).

Minor comments:
Abstract - double-check tenses (Line 9-10)
Line 32 - they = traits evolve
Line 35 - during = following

>> We thank the Referee for these helpful suggestions, which we implemented.

Line 66 - Clarify that MKRF samples are mainland individuals and how many per individuals were sampled per Saturna, Pender and MKRF.
SupplementPage3-Refers to 54 P. maniculatus samples - but the authors describe 28 Saturna, 9 Pender and 28 mainland - details on sample numbers for different genomic analysis need to be clarified

>> We apologize for the confusion. We isolated genomic DNA from 65 specimens, but then filtered specimens as described. The final dataset for Fig.

1B&C contained 54 specimens, the final dataset for Fig. 1D contained 14 specimens. We clarified this in the manuscript.

Line80 - How many male and female individuals for Saturna and MKRF were used to establish colonies?

>> We added this information to the Methods. We imported approximately 20 individuals from each wild population to the lab, where they underwent zoonotic testing and quarantine. Some of these individuals did not breed, and therefore the final founders of our colonies were: 9 females and 9 males from Saturna Island, and 8 females and 3 males from MKRF.

Line 183 - how many island v. mainland individuals were used in this comparison?

>> The sample size is shown in Fig. 2. We used 10 wild-caught and 10 captive-born mainland mice, and 11 wild-caught and 15 captive-born island mice, for a total of 21 wild-caught and 25 captive-born mice.

Line 193 - Re-organization of this sentence phrasing statement as a test-able hypothesis would clarify the question asked here.

>> Thank you. We changed the sentence to read: “We next tested whether island mice were heavier because they are larger (i.e. longer) and/or because they are heavier relative to their body length.” (line 187-188).

Justification for methods is placed in the results oftentime (ie: why X-ray measurements) and the manuscript could be re-organized to justify approach within the method over results section.

>> We agree with the Referee that the Results section contained some detail about the methods used, and we moved several Results sections to the Methods, for example, L107-108, L110-114, L130-132.

**Referee: 2**

This study investigated the genetic and plastic underpinnings of morphology and behavior between island and mainland deer mice. The authors have done an impressive amount of work to collect these data, in both wild and lab-raised animals. Overall, this is a nice paper characterizing differences between island and mainland deer mice.

>> We thank the Referee for his/her support of our study.

The majority of the findings have already been shown in previous studies regarding these same populations of deer mice. In particular, population differences in body size have been shown previously in a large number of studies, and the authors highlight these similarities in the first two paragraphs of the ‘Discussion’ section (lines 368-398). This current study provides further characterization of body size differences between island and mainland deer mice by showing that island mice are born heavier, with the greatest growth occurring between birth and weaning, and that a maternal effect contributes to these size differences. This study also demonstrates that body size itself has a genetic basis, which has been shown previously in other mouse populations, including island mice (Gray et al., 2015; Roth et al., 1986; Parmenter et al., 2016; Phifer-Rixey et al., 2018). Lastly, it was previously shown that these island and mainland mice are genetically distinct with little to no ongoing gene flow between island and mainland or among island populations (Redfield, 1975). Thus, even with more fine-scaled measurements and characterization, this portion of the study provides very little novel insights regarding body size evolution in island and mainland populations of mice.

>> We agree with the Referee that some aspects of our results are consistent with results from previous studies (which were, of course, the motivation to conduct this more in-depth study!). Nonetheless, we feel we need to show some data from the precise populations we are studying and with our methods and assays in order for the reader to be able to make direct comparisons between, for example, our wild versus laboratory-raised mice. This is a necessary foundation for our more mechanistic findings (and a comparison to our behavioral work).

In addition, in many cases, we generated more or better data to support our claims; for example, previous work suggests minimal gene flow between mainland and island populations, but this was based on electrophoresis from only a single protein marker and no estimates of divergence time were made. By contrast, we sampled thousands of SNP markers from across the genome and then use these genomic data to make well supported estimates of both gene flow and divergence times.

Moreover, we wanted to note that we have been very careful and honest about citing the work of others (all the papers noted above are cited in our paper, with the exception of the new Phifer-Rixey house mouse study that doesn’t sample islands).

Nonetheless, to address this concern, we dramatically streamlined these sections where possible (they now make up only two very short paragraphs in the Results), moved a figure panel to the supplement, and used language suggested by the Associate Editor to be even more clear where our results are confirmatory or consistent with previous studies.

The second half of this study describes aggression differences between island and mainland mice. Previous studies found similar results (lines 307-309 and line 404): wild-caught island deer mice show reduced territorial aggression compared to wild-caught mainland deer mice. The novel aspect of this study is measuring aggression in lab-reared mice to control for environmental effects on behavior. The authors demonstrate that wild-caught mice behave differently than lab-reared animals, implying that some aspects of aggression differences are driven by environmental effects. The authors do a great job describing the various factors that may be underlying wild aggression differences in these two populations (lines 403-416).

>> We thank the Referee for his/her support of our behavioral work.

However, it is difficult to accurately disentangle the underlying mechanisms behind the differences seen in aggression between all three experiments (wild captive, short-term captive, long-term captive), as each experiment induced different levels of aggression.

>> Thank you for this comment, however, we feel the opposite is true: Aggression levels differ between the experiments *because* we successfully disentangled some of the underlying mechanisms. In other words, despite being able to induce high levels of aggression, we still did not see differences between island and mainland mice when we minimized environmental effects (i.e. in lab-reared mice).

For example, although differences in aggression in wild-captive mice cannot be explained by reproductive experience (Fig. S10; Fig. 4), reproductive experience does play a role in aggression behavior in lab-reared mice (Fig. 5 and S11; lines 339-342). Perhaps, instead of environmental effects, past reproductive experience is why Fig. 4 (wild) and Fig. 5 (lab-reared) show different results?

>> We demonstrate that reproductive experience is important for inducing aggression in these mice (Fig. 5, S12). Most wild-caught mice (Fig. 4) and all captive-born mice in the high-aggression experiment (Fig. 5C-D) had reproductive experience at the time they were tested. And yet, mainland-island mice differ in aggression levels in the wild-caught experiment, but not in the long-term captive experiment. Therefore, past reproductive experience alone cannot explain the aggression differences in the wild-caught mice.

Moreover, the authors specifically induced higher aggression levels in captive-born mice (Fig. 5C & 5D) compared to the behavioral assay induced on wild-captive mice. If the wild-captive mice were also exposed to a high aggression assay (like in Fig. 5C), would they have shown similar results as in Fig. 5D?

>> The main difference between the wild-captive experiment in Fig. 4 and the high-aggression assay in Fig. 5C-D (besides the fact that mice were wild-born in one and captive-born in the other) is that mice in the high-aggression assay were tested immediately after copulation. The Referee's question can therefore be re-phrased: Would wild mice show increased aggression if they had been tested immediately after copulation? We have not done this experiment.

However, we have disentangled the effect of recent copulation vs. more distant reproductive experience on aggression in captive mice: Compare the right column of S12B to Fig. 5D - these experiments are largely comparable besides the fact that mice were tested after copulation in the latter. Wrestling duration is similar in these experiments, suggesting that reproductive experience, more so than recent copulation, is the primary driver of aggression in both island and mainland captive mice. Thus, we suggest the answer is "no"; recent copulation would not dramatically change the behavior of wild mice.

More importantly, however, why would it matter biologically if the answer would be "yes"? It seems to us the point here is simply that wild island mice show less aggression than mainland mice *in some experimental condition*. To conclude that they show differences in aggression does not require that they must be less aggressive across all experimental conditions.

Or if the authors expose captive-bred mice to the same assay the wild-captive mice underwent, would the authors see similar levels of aggression in both wildcaught and lab-reared mice?

>> We have done this experiment. Captive-born mice that were re-tested after they had sired a litter were essentially exposed to the same biologically relevant conditions as the wild-caught mice: prolonged co-housing with a female and reproductive experience of siring and raising a litter. This is a comparison of wrestling duration in the right column of S12B to wrestling duration in Fig. 4. While we see similar levels of aggression in captive mice (S12B), mainland mice are more aggressive in the wild-caught experiment (Fig. 4), suggesting that environmental factors specific to the wild mice must explain the difference.

Overall, these assay differences make it difficult to make a direct comparison between wild and lab-reared mice, and to overall disentangle what is plastic and what is an artifact of assay design/differences.

>> We have addressed the Referee's specific concerns above. However, we would like to offer additional explanations of our experiments below that (we hope) will reassure the Referee that these experiments are well controlled and comparable.

It is perhaps relevant to point out first that the behavioral assay itself (arena setup, time of day, experimenter, testing room, habituation times, etc.) was always the same across experiments.

What was different between experiments is only the individual experience of mice before the behavioral testing. The principal differences in experience can be summarized as follows:

Experiment 1. Wild mice in captivity.
Co-housing and reproductive experience for most mice (Fig. 4, S11A).

Experiment 2. Low aggression of captive-born mice.
Co-housing, but little reproductive experience (Fig. 5A&B, S12A).

Experiment 3. Re-tested mice from experiment 2.
Co-housing and reproductive experience (S12B).

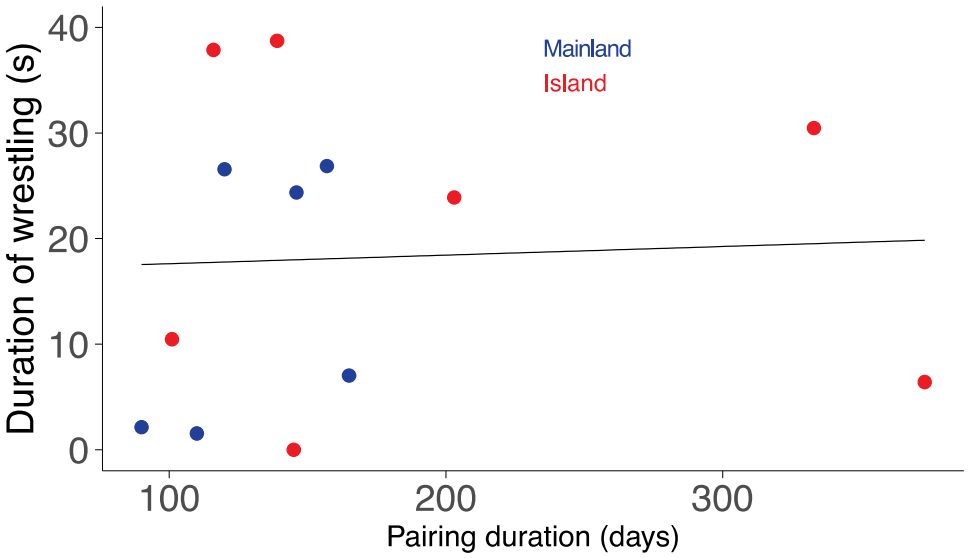
Experiment 4. High aggression of captive-born mice.
Co-housing, reproductive experience, and testing immediately after copulation, i.e. the male is aware of the female's receptivity (Fig. 5C&D).

In other words, biologically meaningful differences in experience revolve largely around the presence/absence of mating experience.

The Referee's questions above and further below suggest that concerns about how comparable these experiments are may in part arise because of the variation in pairing duration and age of mice between experiments. Indeed, our terminology in the manuscript and figure design (e.g. in Fig. 5: "short-term" vs. "long-term" experiments and labels that provide pairing duration in the different experiments) may have suggested that pairing duration influences behavior.

However, while longer pairing up to a point certainly increases the probability of mating in deer mice (see S12A), our experiment makes a case that it is really the presence/absence of reproductive experience itself that is biologically relevant and increases aggression.

To illustrate this point explicitly, we re-analyzed the dataset in Fig. 5C-D. All mice in this experiment had reproductive experience, yet they had been paired for substantially different durations by the time of testing, ranging from 3-14 months (we noticed during re-analysis that we described pairing duration in Fig. 5 as "~6-12 months", and have changed this to provide a more precise min-max range of 3-14). As expected, re-analysis shows that there is no effect of pairing duration (and by extension, age - because pairing duration and age are strongly correlated) on aggression levels:




Thus, while there are additional differences between the experiments, we are convinced the biologically meaningful differences listed above really explain most of the behavior differences.

**Together, we believe our complementary set of experiments demonstrates that we can "titrate" aggression in all captive-born mice from low to high, and yet still find no evidence for differences between the island and mainland populations. We believe this (1) argues against a measurable, genetically-based, invariably expressed difference in behavior, and (2) suggests that the behavior differences in wild-caught mice (island v mainland) likely arise largely from an environmental component.**

Due to these discrepancies, I suggest the authors omit sentences 327-328: "…both island and mainland lab-reared animals behaved similarly to the less aggressive wild-caught island mice.", and sentences 363-364: "…the aggression differences observed in wild-caught mice are driven, at least in large part, by environmental effects". The authors do not have the power to make the comparison between lab-reared animals and wild-caught island mice at this time. All the authors can confidently say is that different behavioral assays show similar or different results.

>> We respectfully disagree with the Referee. L327-328 (L291-293 in the revised MS) only compares the behavior of mice in one experiment to the behavior of mice in another experiment, without making any claims as to why that might be the case. And we would like to retain L363-364 (L325-327) for reasons listed above – we believe our set of experiments demonstrates that the behavior differences in wild mice have a strong environmental component.

In the rest of the discussion (lines 418-446), the authors attempt to frame the behavioral differences seen between experiments in an evolutionary context.

There are some good hypotheses, but overall, it is all very speculative and adds very little to the paper. I suggest omitting this entire discussion. Below are some major comments regarding this portion of the discussion section. Although it is likely that some aspects of plasticity are at play here, it is completely speculative to try and decipher the evolutionary role territorial plasticity has played in island and mainland deer mice, as the results do not give much support to these speculations:

>> We agree with the Referee that these parts (L418-446) of the discussion are speculative. Therefore, we removed these two paragraphs and include only one sentence (clearly labeled as a speculation) about the evolution of morphological and behavior in these island mice (L394-399).

1. Lines 425-428: The authors state that the island population has the capacity to be plastic in territorial behavior, and this ability allowed them to colonize the island. However, it appears that the mainland population also shows behavior plasticity (e.g. compare 'chasing' and 'pindown' behaviors between Fig. 4C and Fig. 5). In fact, in lines 326-328, the authors state that wild mainland mice show higher levels of aggression than labreared mainland. Therefore, it seems that both island and mainland populations have the ability to be plastic in territorial aggression. But again, it is difficult to decipher what is actually plastic and what is an artifact of assay differences between each experiment. If it really is plasticity, it suggests that ancestral plasticity (i.e. plasticity in the mainland population) is still present in both mainland and island populations, and this 'ancestral' plasticity in territorial behavior likely allowed the mice to colonize islands. And as the authors suggest in lines 456-460, perhaps selection on aggression is too weak to see any heritable differences between the two populations.

>> We thank the Referee for pointing this out and agree with his/her statement. We have no reason to assume there is a genetically-based difference in the ability for plasticity between island and mainland mice. The most parsimonious assumption is that behavior is controlled by ancestral plasticity in both populations. We have removed this discussion as suggested in the preceding comment.

2. Lines 440-446 (regarding genetic assimilation) are highly speculative. Both island and mainland populations of deer mice show morphological and behavioral plasticity. In fact, captive, mainland deer mice are able to increase in body size similar to that of wild, island deer mice (Fig. 2).

>> As suggested above, we removed this paragraph.

How are the authors able to decipher that this morphological plasticity proceeds the behavioral plasticity seen between the two populations? Especially since both morphological and behavioral plastic responses appear to go in the same

direction as the evolved, island response (i.e. bigger body size and less aggressive).

>> We agree the statement that behavioral plasticity preceded morphological evolution is speculative, although we note that the literature has repeatedly emphasized that behavior often responds more quickly to environmental perturbation than morphology (e.g. Foster 2013, Losos et al 2004, Price et al 2008, Yeh et al 2004). We have removed this paragraph from the discussion as noted above, and instead have only one sentence related to this point (L394-399).

Additional comments for the authors:

3. Why did the authors choose to only screen males in the behavior assays, and not females? It has been previously shown that female aggression (i.e. aggression towards juveniles, specifically) likely plays an important role in these two populations (Halpin 1981). Just curious as to why the authors did not assay female aggression.

>> We agree with the Referee that it would have been interesting to assess female aggression as well. However, because of time and resource constraints this was beyond the scope of the current study.

4. The methods section needs more information. A large portion of the methods are included in the results, making it hard for the reader to decipher which parts of each experiment are the same and/or different. I suggest the authors elaborate on the methods and move aspects of the results to the methods section (lines 271-279, 315-316, and 343-351). In addition, the authors need to state the ages of the lab-reared mice used in the long-term captive behavioral assays (Fig. 5C and 5D). Were the mice similar in age (approximately) to the wild-captive mice (Fig. 4)?

>> We agree and moved several sentences in the Results sections to the Methods, for example, (in the revised MS version) L107-108, L110-114, L130-132. We also added the age of the lab-reared mice to the Methods section (L119). The lab-reared mice were younger than the wild-captive mice; however as shown above, aggression does not change with age/pairing duration and captive island/mainland mice of advanced age still show high levels of aggression without differences between the populations.

# Appendix B

## Associate Editor

I find the authors have done a thorough job of addressing issues raised and/or justifying why changes were not made.

>> We thank the Associate Editor for his/her comment.

I rather like the Referee's suggestion for testing for retained differences in aggression and suggest this is done unless there are good reasons for not doing so.

>> We agree and have added a new supplemental figure (Fig S14) as suggested by the Referee.

Also, while I agree that inter-year variation is unlikely to create trends, I do think inter-year variation is likely present. One solution would be to fit year as a random factor. This might well reveal stronger trends by removing a chunk of the error variance.

>> We tested fitting year as a random factor, using a linear mixed effects model. As expected, the effect of land type (island vs. mainland) on weight is highly significant (t = 9.074, $P < 2e\text{-}16$). However, the trend is not stronger than the result we obtained with the Kruskal-Wallis test (Chi-squared = 185.25, $P < 2.2e\text{-}16$; reported in the paper as "$P < 0.0001$" for simplicity). Thus, we did not change the statistics in the manuscript.

## Referee 2

The authors have done a great job addressing my concerns associated with the original manuscript. I especially like what the authors have done in streamlining the body size results - they clearly state where their results are confirmatory or consistent with previous studies (e.g. L319-330). I find the latter half of the body size results (i.e. size differences at birth and an island-maternal genetic effect) the most compelling and exciting aspect of this paper. The genetic and divergence analyses are also solid and confidently show that the island and mainland populations are distinct.

>> We thank the Referee for his/her comment.

After reading through the authors' responses regarding aggression plasticity, it is apparent that the authors and I are approaching plasticity differently. The authors are most interested in differences between populations (or lack thereof) in some experimental condition to infer plasticity, while I'm asking about differences within populations across wild and captive experiments to infer plasticity. The latter is typically plotted as a reaction norm, which the authors have done for body size differences within and between populations (i.e. Fig. 2, Fig. S4, and Fig. S5). For

example, the authors found that while body weight and fat/lean mass are plastic traits, body length is not, suggesting that environmental conditions have little effect on body length (discussed in L182-187). I was hoping to see a similar reaction norm-analysis for aggression behavior, but as I mentioned in my original comments, this is difficult to assess since the authors induced different levels of aggression between the wild-caught mice and the lab-reared experiments.

However, the authors have pointed out where wild-caught mice (i.e. experiment #1) and lab-reared mice (i.e. experiment #3) are comparable. Specifically, Fig. 4D and the right column of Fig. S12B are largely comparable (i.e. prolonged co-housing with a female and reproductive experience of siring and raising a litter; the only difference is environment). I would like to see these data side-by-side so that readers can easily assess aggression plasticity as a reaction norm (just like how the authors plotted Fig. 2). This new figure can be a supplemental figure. The authors will need to run a linear fixed effects model, with a strain by origin interaction to test for plasticity within populations (again just like they did for Fig. 2). It would be great if the authors could also include the other two behaviors for the lab-reared mice (i.e. Fig. S12B but with chasing and pindown duration) - especially pindown duration as this behavior was different between populations in wild mice. This analysis will allow one to confidently assess which behaviors are plastic *within AND between* populations across environments.

As it's plotted now (pre-statistical analysis), captive mice seem to show similar levels of aggression (Fig. S12B). If this result holds up, it is in contrast to what is seen with wildcaught mice (Fig. 4), suggesting that environmental factors likely explain the aggression difference seen in wild mice.

>> We thank the Referee for his/her comment, and agree that a reaction norm plot of aggression data would be helpful. As suggested, we added a supplemental figure (Fig. S14) that plots the data from Fig 4C and the right column of S12B for all three high intensity aggression behaviors (wrestling, chasing, pindown) side by side, and used a linear fixed effects model with a strain by origin interaction to test for statistical differences. As expected, these analyses demonstrate no difference in aggressiveness between wild-caught and captive-born mainland mice. By contrast, wild-caught island mice are less aggressive than captive-born island mice.

These results (experiment #3) should be stated at L266 – directly after experiment #1 (Fig. 4) and before experiment #2 (Fig. 5A-B). After analyzing experiment #1 and experiment #3 together (i.e. Fig. 4 with Fig. S12B), I think the authors should then transition to experiments #2 and #4 (Fig. 5), where aggression levels are titrated from low- to high-aggression. These sets of experiments were performed to ask if variation in reproductive experience can recapitulate aggression differences seen in the wild-caught mice. I think these experiments are excellent and demonstrate that although reproductive experience is important for inducing aggression in these mice, reproductive

experience alone cannot explain the aggression differences seen between wild-caught island and mainland populations. I suggest the authors combine these two results sections (L267-314) under one heading, something like, “Captive-bred mice do not differ in aggression levels, regardless of reproductive experience”.

After these revisions, I agree with the authors that their study:
(1) argues against a measurable, genetically based, invariably expressed difference in behavior in island mice, and
(2) suggests that the behavior differences in wild-caught mice (island v mainland) likely arise from an environmental component.

>> As requested by the Referee, we have combined the two results sections (L267-314) under one heading, "Captive-born mice do not differ in aggression levels, regardless of reproductive experience". However, we have not implemented the suggested re-arrangement of results sections, i.e. moving the discussion of Fig. S12B and Fig. 4 (which now is the new supplemental figure S14) before Fig. 5A&B. Instead, we have added a sentence at the end of the results section about Fig S14, where it further emphasizes the point made throughout, i.e. that reproductive experience is crucial to induce aggression, but that differences in reproductive experience cannot explain differences in aggressiveness, and that instead other environmental factors likely contribute to these aggression differences.

We have not implemented this suggestion because (1) the experiment in Fig. S12B follows logically from the experiment in Fig. 5A&B because it re-tests the animals in this experiment, and thus Fig S12B would be difficult to discuss without having previously discussed the experiment in Fig. 5A&B; and (2) to make the point that the experiments in Fig. S14 are largely equivalent and thus comparable, it is necessary to first introduce and discuss the relevance of reproductive experience (i.e. Fig. 5 and S12) to the reader.

Additional Comments:
L112-121: The authors only describe two experiments (short-term pairing and long-term pairing) when there should be three experiments. The following lines are included in the supplemental methods (pg. 5) and I suggest moving them to the main manuscript:

“For one set of experiments, wild-caught resident males were paired with a female for several months. For a second set of experiments, 9- to 13-week-old lab-reared residents (non-breeding) males were isolated with a female for one week. For a third set of experiments, 5-16-month old lab-reared resident (breeding) males had been paired with a female for several months.”

>> We have implemented this suggestion.

L117: omit the word “also”

>> We have implemented this suggestion.

L130-132: The statistical analyses paragraph of the methods section is too brief. What is stated currently mostly applies to the body size results. Since the behavioral assays are complicated, I suggest the authors elaborate on the statistical analyses used for the aggression experiments by moving some of the supplemental methods to the main manuscript.

>> We agree that the description of the statistical analyses is brief; it basically only lists the R packages we used (however, all packages are listed and it is not biased towards the body size results). We moved all specific details of the statistical analysis to the Supplement because of the space constraint of 10 pages (we don’t even have 10 words to spare!). Unfortunately, there is no simple way to elaborate on this description without biasing some aspects of the results over others.

L223: state ‘male hybrids’ instead of ‘male’

>> We have implemented this suggestion.

L266: Before stating the results of experiment #2, I suggest the authors state the results of the comparison between wild-caught mice (Fig. 4) and captive-born mice (experiment #3, Fig. S12B). Again, this analysis can be shown in a supplemental figure.

>> As discussed above, we chose to add this discussion at the very end of the results section to not disrupt the logical flow of the previous experiments.

L269-270: The low-aggression experiment does not accurately fulfill this goal. Instead, experiment #3 answers this question. I suggest the authors use this sentence to introduce the results of experiment #3, which should come before the low-aggression experiment results.

>> We respectfully note that L269-270 intends to shift the attention of the reader away from wild-caught mice to lab-born mice to introduce this new paradigm. It does not intend to say that the low-aggression experiment alone answers this question. We attempted to make this clear by starting the next sentence with the word "initially", i.e. implying that there will be more experiments subsequently to provide a more complete answer to this question.

In my opinion, the goals of the low-aggression and high-aggression experiments are complementary since together they demonstrate that (1) reproductive experience is important for inducing aggression in these mice, yet (2) reproductive experience alone cannot explain the aggression differences seen

between wild-caught island and mainland populations, since captive bred mice remain indistinguishable (discussed in L350-352). These two experiments (Fig. 5) can be stated under a single section, titled something like, “Captive-bred mice do not differ in aggression levels, regardless of reproductive experience”.

>> We agree and have implemented this suggestion.

L291-295: I think the authors meant to reference Fig. S12B and not S11B here. Again, these results (experiment #3) should be compared to experiment #1 (Fig. 4) to accurately test if behavioral differences observed in wild mice are retained in a controlled laboratory environment over generations. Again, this analysis can be presented as a supplemental figure and should come directly after the wild-caught experiment (L266).

>> We have corrected this.

L354-365: Another explanation besides population density may be the role of female aggression. As I mentioned before, female aggression has been shown to play an important role in these two Peromyscus populations (Halpin, 1981). It would be worthwhile for the authors to briefly discuss the possible role of female vs. male aggression in light of their results.

>> We would love to discuss the possible role of female vs. male aggression, however because of the space constraints, this was not possible to include in the discussion.