

## Life-history traits and habitat availability shape genomic diversity in birds: implications for conservation

Anna Brüniche-Olsen, Kenneth F. Kellner, Jerrold L. Belant and J. Andrew DeWoody

### Article citation details

*Proc. R. Soc. B* **288**: 20211441.

<http://dx.doi.org/10.1098/rspb.2021.1441>

### Review timeline

Original submission: 11 March 2021  
1st revised submission: 24 June 2021  
2nd revised submission: 11 August 2021  
3rd revised submission: 30 September 2021  
Final acceptance: 6 October 2021

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

## Review History

### RSPB-2021-0602.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?**

Acceptable

**Is the length of the paper justified?**

Yes

**Should the paper be seen by a specialist statistical reviewer?**

No

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

Yes

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

**Is it accessible?**

No

**Is it clear?**

N/A

**Is it adequate?**

N/A

**Do you have any ethical concerns with this paper?**

No

### **Comments to the Author**

In their manuscript "Life history traits and habitat availability shape genomic diversity in birds: implications for conservation" authors Brüniche-Olsen et al. use available genomic sequences for 68 avian species to assess the effects of habitat availability (inferred using ecological niche modelling) and life history traits (extracted from the literature) on genome-wide heterozygosity (H) and on long-term Ne (mean Ne during the last 1 million years based on PSMC estimates). The main aim of the paper is to explore if these genome-wide summary statistics reflect meaningful ecological attributes relevant to conservation efforts.

The study's main findings are: 1) long-term Ne (but not H) is positively correlated with habitat availability. 2) long-term Ne and H show a negative correlation with body size and are also affected by diet. 3) H decreased with increased extinction risk and thus might be used as a leading indicator of demographic trends related to formal IUCN conservation status in birds.

I really appreciate the use of single sample genomic data to orient conservation status as with the surge in the generation of reference genomes, these may provide an important tool for conservation, and thus, exploring which genome-wide summary statistics might be more suitable to guide conservation actions is paramount. I think this manuscript has a lot of potential, but I have a few major and minor comments that I would like the authors to address.

### **MAJOR COMMENTS**

In general Results and Discussion are favoured as separate sections in Proceedings B although, it may be acceptable to merge these two sections. However, the Methods section should come before the Results and Discussion section and thus the Methods text should be removed from the Results and Discussion section. In addition, a background or Introduction section should be included, and this text removed from the Results and Discussion section as well. These changes may not be required by Proceedings B but the authors should try to stick to the general journal layout.

I would like to see written justification of why normal glm rather than pgl analyses were performed. At the very least, an analysis of phylogenetic signal (i.e. estimating lambda) should be performed to show that the variables analysed do not have a strong phylogenetic signal and thus pgl analyses are not required. But, ideally an approach such as the one proposed by Revell 2010

(doi:10.1111/j.2041-210X.2010.00044.x) should be used to take phylogenetic signal into account.

I miss an explanation in the Results and Discussion section about the reasons why the authors chose to show the results (in Figures 1, 2 and 4) of the multiple regression models that also included effects of habitat area, diet, and body mass instead of the results of the univariate models. I would like to see written justification for this choice especially when some of the variables used did not have a significant effect on either H or mean long-term Ne. I would also like to see written justification about why the authors omitted model selection and did not choose to show the results of the preferred model (i.e. based on AIC or similar) for each dependent variable (H and Ne), instead of showing the results from the multivariate model containing all the predictors.

Despite I have to acknowledge I am no expert in the field, I really like the environmental niche modelling approach used in the paper. However, I feel that by averaging suitable habitat area over the past ~1 million years, there is a complete omission of the importance of the variance of this measure, which is a potentially important predictor of fluctuations in Ne. I think the paper could benefit from modelling past Ne as a function of suitable habitat area at the particular moment the measure was taken. If the authors don't think such an analysis is within the focus of their paper, at least, they should consider to also model the variance (or similar measure of change) in Ne as well as the mean and they should consider the inclusion of variance (or similar measure of change) of suitable habitat area as a predictor.

I think the authors should expand on why the results of identification of sex-linked scaffolds using Satsuma2 were inconsistent. Sex chromosomes are generally removed from analyses to infer effective population sizes due to their reduced effective population sizes compared to autosomes. I understand that identifying sex chromosomes in scaffold-level genome assemblies is not an easy task, but if no alternative method is used to remove the sex chromosomes, the authors should at least perform PSMC analyses with and without the sex chromosomes for the chromosome-level genome assemblies included in the study to assess the differences in Ne estimates.

#### MINOR COMMENTS

Abstract: The Royal Society author guidelines state that abstracts should be no more than 200 words, so consider streamlining the abstract. I propose that the sentence from line 20 to 23 could be removed. Also, from The Royal Society author guidelines, abstracts should not contain references, so consider removing the references from the abstract.

General: make sure that spelling is British English throughout the manuscript. For instance, I could see the word analyzed and the word colored a few times instead of analysed and coloured.

Line 26: Consider adding historical to this sentence to make it clear that no current estimates of Ne have been made: "by historical long-term effective population sizes".

Line 31: What do you mean by Ne and H being inversely related to diet? I would not talk about "inversely related" for a categorical variable like diet.

Line 35, 66-67 and 150: Please, try to avoid the use of the term population genomics (or population genetics) as the data presented here is based on single genomes per species. Consider changing it to "genome-wide summary statistics".

Line 46-47: Consider rephrasing to something like: "For each species, we used a single whole genome sequence and its associated whole-genome sequencing data to estimate genome-wide present-day heterozygosity (H) and to reconstruct historical long-term effective population size (Ne)."

Line 50-52: Please, consider changing the focus of the definition of  $N_e$  towards historical long-term  $N_e$  as this is the parameter that is inferred in the paper.

Line 53 and 56: Just a suggestion: consider giving the ranges for  $H$  and Mean  $N_e$  in the text and the mean and SD in parentheses.

Line 53-54: Consider rephrasing to emphasise that are trajectories over time: "Trajectories of changes in  $N_e$  through time for all 68 species were estimated...".

Line 56-58: It would help the reader to add the purpose of the modelling at the end of the sentence.

Line 62-63: The island-continental species distinction is stated here but not explored in the paper. Maybe it is not needed to state it here as it may mislead the reader towards thinking that this distinction will be relevant on the paper.

Line 63: Reword: "long-term conservation concern" (no "concerns").

General: Make sure to capitalise all IUCN Red List categories, as recommended by the IUCN Red List.

Line 68: Include Critically Endangered in the threatened category.

Line 69-70: Add post-hoc Tukey's Honestly Significant Difference tests p-values in parentheses.

Line 70-71: Rephrase here as the linear models presented here use IUCN Red List category as a predictor and not the other way round as suggested here.

Line 71: Soften the language here: "are likely because" or similar instead of "are because".

Line 73: Add a comma: "to mutation rate [17], changes..."

Line 74: specify that this refers to changes in  $N_e$  among species and not changes in  $N_e$  through time.

Line 81: I find the use of the term cryptic species a little bit confusing here. Wouldn't it be more appropriate to use something like species with the Data Deficient category as you discuss in Line 151?

Line 91: I would add a generally or something similar in this sentence: "generally resulting in population declines during...".

Line 126: I suggest removing the "For instance," as these two sentences talk about different things.

Line 132: Consider changing this sentence to something like: We measured correlation between  $N_e$  and  $H$  with body mass and diet for clarity.

Line 139: What do you mean by genomic viability?

Line 176 and 328: I could not access the github link. I guess it is still not public. Please, make sure you make it public.

Line 215-216: Consider expanding a little bit more what do you mean by: "The dataset represent the species available that passed our downstream QC criteria". I guess you mean avian species

and what were the downstream QC criteria? Also change “represent” by “represents”.

Line 235: Remove “the” from “the genome-wide heterozygosity”.

Line 240: Consider adding a brief description of how SNPs were called and filtered for PSMC analyses.

Line 280-281: Does that mean that there may be species with as few as 31 records that have been included? Or did you mean fewer than 3000 records?

Line 312: There is a missing word between “was” and “from”.

## FIGURES

Figure 1: Just a suggestion: I don’t find drawings of individual species that fall into that category very informative. I would change the drawings by the IUCN Red List symbols.

Figure 2: What are the drawings representing? If the species were chosen for a particular reason, it should be explained in the figure caption.

Figure 3: I would suggest adding the normalised area as in SI figures.

SI 3: Consider expanding the caption to make it more informative. Also change “normalized” to “normalised” in all plots.

SI 5: genome and SRA codes are not shown. Please, fill in the table.

SI 6: the column “# sites with no repeats” appears as one order of magnitude smaller than it would be expected. Please, double check the numbers.

## 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?

Excellent

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

Excellent

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

Excellent

### Is the length of the paper justified?

Yes

### Should the paper be seen by a specialist statistical reviewer?

No

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

No

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

**Is it accessible?**

Yes

**Is it clear?**

Yes

**Is it adequate?**

Yes

**Do you have any ethical concerns with this paper?**

No

### **Comments to the Author**

This is an interesting study on the power of whole-genome analysis of single individuals to advance our understanding of important demographic parameters. The authors estimate long-term effective population size ( $N_e$ ) and present heterozygosity ( $H$ ) of 68 bird species from whole-genome data, and explore the relationship between these two parameters and several natural history and conservation variables. They find a positive correlation between the area of suitable habitat and  $N_e$  (but not  $H$ ), confirm a negative correlation between body mass and  $N_e$  and  $H$ , and as expected they find that carnivores had lower  $N_e$  and  $H$  than primary consumers.

The results are not necessarily novel, and most patterns have been demonstrated before with more limited molecular datasets. For example, it is well established by phylogeographic studies that post-glacial expansions into northern latitudes from southern refugia have led to temperate and boreal populations of organisms with large effective population sizes but low genetic diversity. However, it is remarkable that whole-genome sequence data from a single individual can confirm results that until now required the use of hundreds of individuals. The PSMC results also provide information on  $N_e$  for thousands of generations into the past, which can provide interesting associations between population size and geoclimatic events. The number of species used and the taxonomic representation is also impressive.

The link between patterns of  $N_e$  and  $H$  and biodiversity conservation I find less compelling. The negative relationship between genomic  $H$  and extinction risk is interesting, yet the resolution is not high, and critically endangered species are not significantly different from vulnerable species. So the suggestion that genome-wide estimates of  $H$  can be used as indicators of conservation status seems premature and overly optimistic. Surely conservation status can be done more easily (less expensively) with conventional genetic markers? In any case, I would tone down the direct relevance of these results to conservation, perhaps not referring to it in the title.

Regarding the long-term  $N_e$  estimates from PSMC (Figure 3), error in recent periods seems too high for the last 10000 to 100000 years to provide reliable data that can be associated with recent climatic events, at least for species with long generation times. Perhaps the Holocene should be excluded as little can be said from these analyses for such a recent period. Also, the two species selected for this figure are eminently tropical, yet the effects of the historical events used (Holocene, Last Glacial Max, Last Interglacial) had the most severe effects on temperate and boreal species. Perhaps including a temperate/boreal example would be more informative.

For results in Figures 2 and 4, please include in the legend or in the main text the regression  $R^2$  values for all plots.

## Decision letter (RSPB-2021-0602.R0)

04-Jun-2021

Dear Dr BRÛNICHE-OLSEN:

I am writing to inform you that your manuscript RSPB-2021-0602 entitled "Life history traits and habitat availability shape genomic diversity in birds: implications for conservation" 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,  
Dr Daniel Costa  
<mailto:proceedingsb@royalsociety.org>

Associate Editor  
Comments to Author:

I enjoyed reading this manuscript by Brüniche-Olsen and collaborators. I believe the combination of whole-genome level population genomics and ecological data is original. I'd like to congratulate the authors for their work.

I believe the manuscript can be improved in different aspects, mostly the rationale behind some choices. For example, the selection of analytical approaches could be more elaborate (e.g. use of the glm function in the R analyses), or the emphasis on the results from multiple regression models over univariate ones, or the statistics behind the environmental niche modelling.

I'd agree with the referees about the link between these analyses and conservation being weak, so this should be either further elaborate or toned down. I'd also like to encourage the authors to

address all the referees concerns, including the role of variance of the niche modelling, and the high error in the error in recent periods (10k-100k years) in the Ne estimates from PSMC.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)

In their manuscript "Life history traits and habitat availability shape genomic diversity in birds: implications for conservation" authors Brüniche-Olsen et al. use available genomic sequences for 68 avian species to assess the effects of habitat availability (inferred using ecological niche modelling) and life history traits (extracted from the literature) on genome-wide heterozygosity (H) and on long-term Ne (mean Ne during the last 1 million years based on PSMC estimates). The main aim of the paper is to explore if these genome-wide summary statistics reflect meaningful ecological attributes relevant to conservation efforts.

The study's main findings are: 1) long-term Ne (but not H) is positively correlated with habitat availability. 2) long-term Ne and H show a negative correlation with body size and are also affected by diet. 3) H decreased with increased extinction risk and thus might be used as a leading indicator of demographic trends related to formal IUCN conservation status in birds.

I really appreciate the use of single sample genomic data to orient conservation status as with the surge in the generation of reference genomes, these may provide an important tool for conservation, and thus, exploring which genome-wide summary statistics might be more suitable to guide conservation actions is paramount. I think this manuscript has a lot of potential, but I have a few major and minor comments that I would like the authors to address.

#### MAJOR COMMENTS

In general Results and Discussion are favoured as separate sections in Proceedings B although, it may be acceptable to merge these two sections. However, the Methods section should come before the Results and Discussion section and thus the Methods text should be removed from the Results and Discussion section. In addition, a background or Introduction section should be included, and this text removed from the Results and Discussion section as well. These changes may not be required by Proceedings B but the authors should try to stick to the general journal layout.

I would like to see written justification of why normal glm rather than pgl analyses were performed. At the very least, an analysis of phylogenetic signal (i.e. estimating lambda) should be performed to show that the variables analysed do not have a strong phylogenetic signal and thus pgl analyses are not required. But, ideally an approach such as the one proposed by Revell 2010 (doi:10.1111/j.2041-210X.2010.00044.x) should be used to take phylogenetic signal into account.

I miss an explanation in the Results and Discussion section about the reasons why the authors chose to show the results (in Figures 1, 2 and 4) of the multiple regression models that also included effects of habitat area, diet, and body mass instead of the results of the univariate models. I would like to see written justification for this choice especially when some of the variables used did not have a significant effect on either H or mean long-term Ne. I would also like to see written justification about why the authors omitted model selection and did not choose to show the results of the preferred model (i.e. based on AIC or similar) for each dependent variable (H and Ne), instead of showing the results from the multivariate model containing all the predictors.

Despite I have to acknowledge I am no expert in the field, I really like the environmental niche modelling approach used in the paper. However, I feel that by averaging suitable habitat area over the past ~1 million years, there is a complete omission of the importance of the variance of this measure, which is a potentially important predictor of fluctuations in Ne. I think the paper could benefit from modelling past Ne as a function of suitable habitat area at the particular



moment the measure was taken. If the authors don't think such an analysis is within the focus of their paper, at least, they should consider to also model the variance (or similar measure of change) in  $N_e$  as well as the mean and they should consider the inclusion of variance (or similar measure of change) of suitable habitat area as a predictor.

I think the authors should expand on why the results of identification of sex-linked scaffolds using Satsuma2 were inconsistent. Sex chromosomes are generally removed from analyses to infer effective population sizes due to their reduced effective population sizes compared to autosomes. I understand that identifying sex chromosomes in scaffold-level genome assemblies is not an easy task, but if no alternative method is used to remove the sex chromosomes, the authors should at least perform PSMC analyses with and without the sex chromosomes for the chromosome-level genome assemblies included in the study to assess the differences in  $N_e$  estimates.

#### MINOR COMMENTS

Abstract: The Royal Society author guidelines state that abstracts should be no more than 200 words, so consider streamlining the abstract. I propose that the sentence from line 20 to 23 could be removed. Also, from The Royal Society author guidelines, abstracts should not contain references, so consider removing the references from the abstract.

General: make sure that spelling is British English throughout the manuscript. For instance, I could see the word analyzed and the word colored a few times instead of analysed and coloured.

Line 26: Consider adding historical to this sentence to make it clear that no current estimates of  $N_e$  have been made: "by historical long-term effective population sizes".

Line 31: What do you mean by  $N_e$  and  $H$  being inversely related to diet? I would not talk about "inversely related" for a categorical variable like diet.

Line 35, 66-67 and 150: Please, try to avoid the use of the term population genomics (or population genetics) as the data presented here is based on single genomes per species. Consider changing it to "genome-wide summary statistics".

Line 46-47: Consider rephrasing to something like: "For each species, we used a single whole genome sequence and its associated whole-genome sequencing data to estimate genome-wide present-day heterozygosity ( $H$ ) and to reconstruct historical long-term effective population size ( $N_e$ )."

Line 50-52: Please, consider changing the focus of the definition of  $N_e$  towards historical long-term  $N_e$  as this is the parameter that is inferred in the paper.

Line 53 and 56: Just a suggestion: consider giving the ranges for  $H$  and Mean  $N_e$  in the text and the mean and SD in parentheses.

Line 53-54: Consider rephrasing to emphasise that are trajectories over time: "Trajectories of changes in  $N_e$  through time for all 68 species were estimated...".

Line 56-58: It would help the reader to add the purpose of the modelling at the end of the sentence.

Line 62-63: The island-continental species distinction is stated here but not explored in the paper. Maybe it is not needed to state it here as it may mislead the reader towards thinking that this distinction will be relevant on the paper.

Line 63: Reword: "long-term conservation concern" (no "concerns").

General: Make sure to capitalise all IUCN Red List categories, as recommended by the IUCN Red List.

Line 68: Include Critically Endangered in the threatened category.

Line 69-70: Add post-hoc Tukey's Honestly Significant Difference tests p-values in parentheses.

Line 70-71: Rephrase here as the linear models presented here use IUCN Red List category as a predictor and not the other way round as suggested here.

Line 71: Soften the language here: "are likely because" or similar instead of "are because".

Line 73: Add a comma: "to mutation rate [17], changes..."

Line 74: specify that this refers to changes in  $N_e$  among species and not changes in  $N_e$  through time.

Line 81: I find the use of the term cryptic species a little bit confusing here. Wouldn't it be more appropriate to use something like species with the Data Deficient category as you discuss in Line 151?

Line 91: I would add a generally or something similar in this sentence: "generally resulting in population declines during..."

Line 126: I suggest removing the "For instance," as these two sentences talk about different things.

Line 132: Consider changing this sentence to something like: We measured correlation between  $N_e$  and  $H$  with body mass and diet for clarity.

Line 139: What do you mean by genomic viability?

Line 176 and 328: I could not access the github link. I guess it is still not public. Please, make sure you make it public.

Line 215-216: Consider expanding a little bit more what do you mean by: "The dataset represent the species available that passed our downstream QC criteria". I guess you mean avian species and what were the downstream QC criteria? Also change "represent" by "represents".

Line 235: Remove "the" from "the genome-wide heterozygosity".

Line 240: Consider adding a brief description of how SNPs were called and filtered for PSMC analyses.

Line 280-281: Does that mean that there may be species with as few as 31 records that have been included? Or did you mean fewer than 3000 records?

Line 312: There is a missing word between "was" and "from".

## FIGURES

Figure 1: Just a suggestion: I don't find drawings of individual species that fall into that category very informative. I would change the drawings by the IUCN Red List symbols.

Figure 2: What are the drawings representing? If the species were chosen for a particular reason, it should be explained in the figure caption.

Figure 3: I would suggest adding the normalised area as in SI figures.

SI 3: Consider expanding the caption to make it more informative. Also change “normalized” to “normalised” in all plots.

SI 5: genome and SRA codes are not shown. Please, fill in the table.

SI 6: the column “# sites with no repeats” appears as one order of magnitude smaller than it would be expected. Please, double check the numbers.

Referee: 2

Comments to the Author(s)

This is an interesting study on the power of whole-genome analysis of single individuals to advance our understanding of important demographic parameters. The authors estimate long-term effective population size ( $N_e$ ) and present heterozygosity ( $H$ ) of 68 bird species from whole-genome data, and explore the relationship between these two parameters and several natural history and conservation variables. They find a positive correlation between the area of suitable habitat and  $N_e$  (but not  $H$ ), confirm a negative correlation between body mass and  $N_e$  and  $H$ , and as expected they find that carnivores had lower  $N_e$  and  $H$  than primary consumers.

The results are not necessarily novel, and most patterns have been demonstrated before with more limited molecular datasets. For example, it is well established by phylogeographic studies that post-glacial expansions into northern latitudes from southern refugia have led to temperate and boreal populations of organisms with large effective population sizes but low genetic diversity. However, it is remarkable that whole-genome sequence data from a single individual can confirm results that until now required the use of hundreds of individuals. The PSMC results also provide information on  $N_e$  for thousands of generations into the past, which can provide interesting associations between population size and geoclimatic events. The number of species used and the taxonomic representation is also impressive.

The link between patterns of  $N_e$  and  $H$  and biodiversity conservation I find less compelling. The negative relationship between genomic  $H$  and extinction risk is interesting, yet the resolution is not high, and critically endangered species are not significantly different from vulnerable species. So the suggestion that genome-wide estimates of  $H$  can be used as indicators of conservation status seems premature and overly optimistic. Surely conservation status can be done more easily (less expensively) with conventional genetic markers? In any case, I would tone down the direct relevance of these results to conservation, perhaps not referring to it in the title.

Regarding the long-term  $N_e$  estimates from PSMC (Figure 3), error in recent periods seems too high for the last 10000 to 100000 years to provide reliable data that can be associated with recent climatic events, at least for species with long generation times. Perhaps the Holocene should be excluded as little can be said from these analyses for such a recent period. Also, the two species selected for this figure are eminently tropical, yet the effects of the historical events used (Holocene, Last Glacial Max, Last Interglacial) had the most severe effects on temperate and boreal species. Perhaps including a temperate/boreal example would be more informative.

For results in Figures 2 and 4, please include in the legend or in the main text the regression  $R$  squared values for all plots.

## Author's Response to Decision Letter for (RSPB-2021-0602.R0)

See Appendix A.

## RSPB-2021-1441.R0

### Review form: Reviewer 2 (Borja Mila)

#### **Recommendation**

Major revision is needed (please make suggestions in comments)

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

Excellent

#### **General interest: Is the paper of sufficient general interest?**

Excellent

#### **Quality of the paper: Is the overall quality of the paper suitable?**

Excellent

#### **Is the length of the paper justified?**

Yes

#### **Should the paper be seen by a specialist statistical reviewer?**

No

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

No

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

#### **Is it accessible?**

No

#### **Is it clear?**

No

#### **Is it adequate?**

No

#### **Do you have any ethical concerns with this paper?**

No

#### **Comments to the Author**

I appreciate the changes made to the previous version, which I think have improved the manuscript. However, regarding Figure 3, which combines the PSMC plots with maps of suitable habitat (ENM), I think they remain misleading. The Ne estimates for the Holocene (vertical green bar on Fig 3) are not reliable and typically wrong. This is clear in the new figure for the American crow. There is a clear expansion of habitat area in the Holocene following the Last Glacial

Maximum, yet the subsequent increase in  $N_e$  is not detected by PSMC. I think the authors have two options, either remove the Holocene maps from these plots, or show the increase in the variance of  $N_e$  estimates at these recent dates by showing the bootstrap replicates for each estimate, which have been cut on every plot for the last 10000 years (without justification or explanation). As the plot appears right now, the reader must interpret that following the Last Glacial Maximum, there has been an increase in suitable habitat but not an increase in  $N_e$ , which is not true. Adding the bootstrap replicate lines to the entire plot would at least alert the reader of the large error associated with recent estimates of  $N_e$ .

## Review form: Reviewer 3

### Recommendation

Accept with minor revision (please list in comments)

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

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

In their manuscript, Bruñiche-Olsen and colleagues combine whole genome and ecological data to understand the role of biological and environmental factors in determining levels of genetic diversity to inform conservation. I find that their approach is original, and the paper is well written and accessible to a broad readership. I particularly appreciate the use of some of the

genomes available as it shows how much we can infer about a species from the genome sequence of a single individual. As this is the first time I review this manuscript, I think the authors did a good job addressing previous comments from two reviewers. My major complaint is that this version of the manuscript didn't include line numbers. Please find below some minor comments/suggestions.

End of page 4: could you repeat what the covariates are?

Page 6: generally "cryptic species" has a different meaning as it refers to species that haven't been identified as distinct from another one.

Page 7: this reference to a 'newly discovered baleen whale' is not clear. It reads like the baleen whale has been recently discovered. I suggest replacing with 'a previously undescribed baleen whale species (Balaenoptera spp.)'

Page 8: 'The myriad species' is surely an overstatement when referring to 68 species.

Page 8: I wouldn't consider diet a life history trait per se, but perhaps a proxy for a suite of life history traits. Additionally, aren't body mass and diet highly correlated?

Fig.2 Missing arrow associated with species on the left in the left side panel. Also, please add species depicted in plots in the figure caption.

Page 16: why was ANGSD used for estimating heterozygosity and how? The depth of the data used here should be high for all these species and the SFS for a single individual doesn't really help with estimating heterozygosity in this case. Was heterozygosity the sum of variant sites divided by genome size or by the actual proportion of the genome covered by high quality data? Did you apply some high and low coverage filters?

## Decision letter (RSPB-2021-1441.R0)

05-Aug-2021

Dear Dr BRÜNICH-OLSEN:

Your manuscript has now been peer reviewed and the reviews have been assessed by an Associate Editor. The reviewers' comments (not including confidential comments to the Editor) and the comments from the Associate Editor are included at the end of this email for your reference. As you will see, the reviewers and the Editors have 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=\(Document not available\)](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,  
Dr Daniel Costa  
mailto: proceedingsb@royalsociety.org

Reviewer(s)' Comments to Author:

Referee: 2

Comments to the Author(s).

I appreciate the changes made to the previous version, which I think have improved the manuscript. However, regarding Figure 3, which combines the PSMC plots with maps of suitable habitat (ENM), I think they remain misleading. The  $N_e$  estimates for the Holocene (vertical green bar on Fig 3) are not reliable and typically wrong. This is clear in the new figure for the American crow. There is a clear expansion of habitat area in the Holocene following the Last Glacial Maximum, yet the subsequent increase in  $N_e$  is not detected by PSMC. I think the authors have two options, either remove the Holocene maps from these plots, or show the increase in the variance of  $N_e$  estimates at these recent dates by showing the bootstrap replicates for each estimate, which have been cut on every plot for the last 10000 years (without justification or explanation). As the plot appears right now, the reader must interpret that following the Last Glacial Maximum, there has been an increase in suitable habitat but not an increase in  $N_e$ , which is not true. Adding the bootstrap replicate lines to the entire plot would at least alert the reader of the large error associated with recent estimates of  $N_e$ .

Referee: 3

Comments to the Author(s).

In their manuscript, Brüniche-Olsen and colleagues combine whole genome and ecological data to understand the role of biological and environmental factors in determining levels of genetic diversity to inform conservation. I find that their approach is original, and the paper is well written and accessible to a broad readership. I particularly appreciate the use of some of the genomes available as it shows how much we can infer about a species from the genome sequence of a single individual. As this is the first time I review this manuscript, I think the authors did a good job addressing previous comments from two reviewers. My major complaint is that this version of the manuscript didn't include line numbers. Please find below some minor comments/suggestions.

End of page 4: could you repeat what the covariates are?

Page 6: generally "cryptic species" has a different meaning as it refers to species that haven't been identified as distinct from another one.

Page 7: this reference to a 'newly discovered baleen whale' is not clear. It reads like the baleen whale has been recently discovered. I suggest replacing with 'a previously undescribed baleen whale species (*Balaenoptera* spp.)'

Page 8: 'The myriad species' is surely an overstatement when referring to 68 species.

Page 8: I wouldn't consider diet a life history trait per se, but perhaps a proxy for a suite of life history traits. Additionally, aren't body mass and diet highly correlated?

Fig.2 Missing arrow associated with species on the left in the left side panel. Also, please add species depicted in plots in the figure caption.

Page 16: why was ANGSD used for estimating heterozygosity and how? The depth of the data used here should be high for all these species and the SFS for a single individual doesn't really help with estimating heterozygosity in this case. Was heterozygosity the sum of variant sites



divided by genome size or by the actual proportion of the genome covered by high quality data?  
Did you apply some high and low coverage filters?

## Author's Response to Decision Letter for (RSPB-2021-1441.R0)

See Appendix B.

## RSPB-2021-1441.R1 (Revision)

### 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?**

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?**

No

#### **Is it clear?**

N/A

#### **Is it adequate?**

N/A

#### **Do you have any ethical concerns with this paper?**

No

#### **Comments to the Author**

I appreciate the authors' reply to my comment, but I am afraid their arguments are insufficient and the main problem remains. They explain in the methods that "Very recent estimates of Ne

from PSMC tend to be noisy thus following Leroy et al. [73], we excluded the four most recent time points of  $N_e$  for each species when calculating the mean estimate to increase reliability." However, Leroy et al. (2021) removed those four recent points because they were interested in obtaining an average  $N_e$  value for the last million years, and were not specifically interested in the most recent few thousand years, which is where PSMC is unreliable. But here the situation is clearly different, as the PSMC is used specifically to infer  $N_e$  during the Holocene, and establishing a direct comparison with the extent of habitat availability that is actually depicted in a map within the same figure. If the authors remove the four most recent time intervals, then there is actually no data for the Holocene, and that is why the plot line is flat on the left end on all species. In my last review I requested that bootstrap lines be added to all time intervals to reveal variance, but it is now clear that the data were removed, so the problem is more serious (and the bootstrap lines are meaningless). Ideally the dataset would include several genomes per species, instead of just one, so that other approaches could be used to increase the precision of  $N_e$  estimates in recent time periods (msmc, momi2, etc.). But given that only one genome is available per species, I think the plots should either show the  $N_e$  values for the recent time intervals (revealing noisy variation), or exclude the Holocene maps in the lower panels, as no reliable data seem to be available for that period.

## Decision letter (RSPB-2021-1441.R1)

17-Sep-2021

Dear Dr BRÜNICHÉ-OLSEN:

Your manuscript has now been peer reviewed and the reviews have been assessed by an Associate Editor. The reviewers' comments (not including confidential comments to the Editor) and the comments from the Associate Editor are included at the end of this email for your reference. As you will see, the reviewers and the Editors have 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=\(Document not available\)](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,  
Dr Daniel Costa  
Editor, Proceedings B  
mailto: proceedingsb@royalsociety.org

Associate Editor  
Board Member: 1  
Comments to Author:

I would like to say thanks to the authors for their efforts in improving the paper. Most of the previous concerns have been addressed. However, the issue raised by the referee about the impact of removing effective population size datapoints for the Holocene is significant and should be addressed before publication.

Reviewer(s)' Comments to Author:  
Referee: 2

Comments to the Author(s)

I appreciate the authors' reply to my comment, but I am afraid their arguments are insufficient and the main problem remains. They explain in the methods that "Very recent estimates of  $N_e$  from PSMC tend to be noisy thus following Leroy et al. [73], we excluded the four most recent time points of  $N_e$  for each species when calculating the mean estimate to increase reliability." However, Leroy et al. (2021) removed those four recent points because they were interested in obtaining an average  $N_e$  value for the last million years, and were not specifically interested in the most recent few thousand years, which is where PSMC is unreliable. But here the situation is clearly different, as the PSMC is used specifically to infer  $N_e$  during the Holocene, and establishing a direct comparison with the extent of habitat availability that is actually depicted in a map within the same figure. If the authors remove the four most recent time intervals, then there is actually no data for the Holocene, and that is why the plot line is flat on the left end on all species. In my last review I requested that bootstrap lines be added to all time intervals to reveal variance, but it is now clear that the data were removed, so the problem is more serious (and the bootstrap lines are meaningless). Ideally the dataset would include several genomes per species, instead of just one, so that other approaches could be used to increase the precision of  $N_e$  estimates in recent time periods (msmc, momi2, etc.). But given that only one genome is available per species, I think the plots should either show the  $N_e$  values for the recent time intervals (revealing noisy variation), or exclude the Holocene maps in the lower panels, as no reliable data seem to be available for that period.

## Author's Response to Decision Letter for (RSPB-2021-1441.R1)

See Appendix C.

## Decision letter (RSPB-2021-1441.R2)

06-Oct-2021

Dear Dr Brüniche-Olsen

I am pleased to inform you that your manuscript entitled "Life history traits and habitat availability shape genomic diversity in birds: implications for conservation" 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](mailto: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,

Dr Daniel Costa

Editor, Proceedings B

<mailto:proceedingsb@royalsociety.org>

# Appendix A

Dear Editor,

Thank you for sending the reviewer comments. We have considered each comment and made changes to the manuscript. We hope that this revised version of our manuscript reads clearer and is now acceptable for publication in Proceedings B.

Sincerely,  
Anna Brüniche-Olsen

Associate Editor

Comments to Author:

I enjoyed reading this manuscript by Brüniche-Olsen and collaborators. I believe the combination of whole-genome level population genomics and ecological data is original. I'd like to congratulate the authors for their work.

I believe the manuscript can be improved in different aspects, mostly the rationale behind some choices. For example, the selection of analytical approaches could be more elaborate (e.g. use of the glm function in the R analyses), or the emphasis on the results from multiple regression models over univariate ones, or the statistics behind the environmental niche modelling.

I'd agree with the referees about the link between these analyses and conservation being weak, so this should be either further elaborate or toned down. I'd also like to encourage the authors to address all the referees concerns, including the role of variance of the niche modelling, and the high error in the error in recent periods (10k-100k years) in the Ne estimates from PSMC.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)

In their manuscript "Life history traits and habitat availability shape genomic diversity in birds: implications for conservation" authors Brüniche-Olsen et al. use available genomic sequences for 68 avian species to assess the effects of habitat availability (inferred using ecological niche modelling) and life history traits (extracted from the literature) on genome-wide heterozygosity (H) and on long-term Ne (mean Ne during the last 1 million years based on PSMC estimates). The main aim of the paper is to explore if these genome-wide summary statistics reflect meaningful ecological attributes relevant to conservation efforts.

The study's main findings are: 1) long-term Ne (but not H) is positively correlated with habitat availability. 2) long-term Ne and H show a negative correlation with body size and are also affected by diet. 3) H decreased with increased extinction risk and thus might be used as a leading indicator of demographic trends related to formal IUCN conservation status in birds.

I really appreciate the use of single sample genomic data to orient conservation status as with the surge in the generation of reference genomes, these may provide an important tool for conservation, and thus, exploring which genome-wide summary statistics might be more suitable to guide conservation actions is paramount. I think this manuscript has a lot of potential, but I have a few major and minor comments that I would like the authors to address.

MAJOR COMMENTS

In general Results and Discussion are favoured as separate sections in Proceedings B although, it may be acceptable to merge these two sections. However, the Methods section should come before the

Results and Discussion section and thus the Methods text should be removed from the Results and Discussion section. In addition, a background or Introduction section should be included, and this text removed from the Results and Discussion section as well. These changes may not be required by Proceedings B but the authors should try to stick to the general journal layout.

**RESPONSE:** We acknowledge the reviewers comment and have included an Introduction section. Proceedings B publish papers where the Results and Discussion sections are combined (see i.e., <https://doi.org/10.1098/rsbl.2010.0638>, <https://doi.org/10.1098/rspb.2020.0690>, <https://doi.org/10.1098/rspb.2014.2777>) and where the Methods come after the Discussion (see i.e., <https://doi.org/10.1098/rspb.2020.0070>, <https://doi.org/10.1098/rspb.2020.2192>, <https://doi.org/10.1098/rspb.2017.1735>). To preserve the flow of the manuscript we have chosen to keep the order of the manuscript as is.

I would like to see written justification of why normal glm rather than pgl analyses were performed. At the very least, an analysis of phylogenetic signal (i.e. estimating lambda) should be performed to show that the variables analysed do not have a strong phylogenetic signal and thus pgl analyses are not required. But, ideally an approach such as the one proposed by Revell 2010 (doi:10.1111/j.2041-210X.2010.00044.x) should be used to take phylogenetic signal into account.

**RESPONSE:** We agree with the reviewer that the analyses would be improved by conducting PGLS. We replaced our standard multiple regression models with PGLS multiple regressions (described in detail in the methods) that incorporated phylogenetic signal. We obtained estimates of lambda > 0 (0.40 for H, 0.55 for Ne), thus we chose to present these PGLS results in the revised manuscript. However, the overall results of the analysis, in terms of the direction and magnitude of the covariate effects, was not greatly changed.

I miss an explanation in the Results and Discussion section about the reasons why the authors chose to show the results (in Figures 1, 2 and 4) of the multiple regression models that also included effects of habitat area, diet, and body mass instead of the results of the univariate models. I would like to see written justification for this choice especially when some of the variables used did not have a significant effect on either H or mean long-term Ne. I would also like to see written justification about why the authors omitted model selection and did not choose to show the results of the preferred model (i.e. based on AIC or similar) for each dependent variable (H and Ne), instead of showing the results from the multivariate model containing all the predictors.

**RESPONSE:** Our objective in this study was simply to determine whether each individual covariate had a significant effect on the two response variables. One approach to answering this, as the reviewer implies, would have been to conduct a series of univariate tests (regressions, t-tests, chi-square, as necessary), one per covariate. However, when testing for the effect of a given covariate, we wanted to make sure to control for the effects of the other covariates. Thus we did not conduct any univariate tests and simply fit a 'global' multiple regression model containing all covariates of interest for each response variable. This allowed us to determine directly if each covariate affected H and/or Ne in a simple but robust manner.

As the reviewer points out, another approach would be to use model selection, perhaps by taking these global models and defining a number of nested models containing some combinations of the covariates, and then ranking the models by AIC. We feel this approach is not necessary for our objectives and would simply add unnecessary complexity. We were not interested in determining the "most parsimonious" model balancing explanatory power and number of parameters (e.g. the maximum AIC). Rather we just wanted to know if each covariate had a significant effect on H and/or Ne or not (while controlling for other covariates), and then presented each of these response-covariate relationships regardless of statistical significance so readers would have all the information. We have added additional justification for this approach in the methods.

Despite I have to acknowledge I am no expert in the field, I really like the environmental niche modelling approach used in the paper. However, I feel that by averaging suitable habitat area over the past ~1 million years, there is a complete omission of the importance of the variance of this measure, which is a potentially important predictor of fluctuations in Ne. I think the paper could benefit from modelling past Ne as a function of suitable habitat area at the particular moment the measure was taken. If the authors don't think such an analysis is within the focus of their paper, at least, they should consider to also model the

variance (or similar measure of change) in Ne as well as the mean and they should consider the inclusion of variance (or similar measure of change) of suitable habitat area as a predictor.

**RESPONSE:** We agree that it would be interesting to model past Ne as a function of suitable habitat at particular points in time; in fact we spent a considerable amount of time trying to do this. The issue is that the relationship between Ne and habitat is likely to be species-specific, and within a given species, we only have 3-4 data points (because we were only able to estimate past habitat during 3-4 time periods given limitations of available climate data). This was simply not enough information to get anything useful from these analyses.

We agree with the reviewer that looking at variance instead of mean for Ne and habitat could be interesting. However, rather than variance we believe that coefficient of variation (CV;  $sd/mean$ ) would be more informative in this case, because e.g. if a species had high mean Ne we would also expect it to have higher variance in Ne (and thus variance provides similar information to the mean). We conducted three additional multiple regression analyses: (1) mean Ne modeled as a function of the same covariates as previously, but with habitat CV replacing mean habitat area; (2) Ne CV modeled as a function of the same covariates; and (3) Ne CV modeled as a function of the same covariates but with habitat CV replacing mean habitat area. Overall we did not find any interesting patterns using these models. Given that these questions were outside our primary objectives and our SI is already very long, we have chosen not to include these results in the main manuscript or the SI. If the reviewer/editor feels strongly that they should be included in the SI we can add them.

I think the authors should expand on why the results of identification of sex-linked scaffolds using Satsuma2 were inconsistent. Sex chromosomes are generally removed from analyses to infer effective population sizes due to their reduced effective population sizes compared to autosomes. I understand that identifying sex chromosomes in scaffold-level genome assemblies is not an easy task, but if no alternative method is used to remove the sex chromosomes, the authors should at least perform PSMC analyses with and without the sex chromosomes for the chromosome-level genome assemblies included in the study to assess the differences in Ne estimates.

**RESPONSE:** We agree that the best approach is to remove sex-linked scaffolds from the analyses but it was not achievable to do consistently for our dataset. Of the 68 bird species included in our study only two species were mapped to chromosome level assemblies. Both of these assemblies were from male individuals and therefore only represented the Z chromosome. We have included more information in the Methods explaining how including sex chromosomes in PSMC may influence the trajectories.

#### MINOR COMMENTS

Abstract: The Royal Society author guidelines state that abstracts should be no more than 200 words, so consider streamlining the abstract. I propose that the sentence from line 20 to 23 could be removed. Also, from The Royal Society author guidelines, abstracts should not contain references, so consider removing the references from the abstract.

**RESPONSE:** We have shortened the Abstract and removed the embedded references.

General: make sure that spelling is British English throughout the manuscript. For instance, I could see the word analyzed and the word colored a few times instead of analysed and coloured.

**RESPONSE:** Done.

Line 26: Consider adding historical to this sentence to make it clear that no current estimates of Ne have been made: "by historical long-term effective population sizes".

**RESPONSE:** Done.

Line 31: What do you mean by Ne and H being inversely related to diet? I would not talk about "inversely related" for a categorical variable like diet.

**RESPONSE:** We agree that this was poorly worded and have removed "inversely" from the sentence.

Line 35, 66-67 and 150: Please, try to avoid the use of the term population genomics (or population



genetics) as the data presented here is based on single genomes per species. Consider changing it to “genome-wide summary statistics”.

**RESPONSE:** Done.

Line 46-47: Consider rephrasing to something like: “For each species, we used a single whole genome sequence and its associated whole-genome sequencing data to estimate genome-wide present-day heterozygosity (H) and to reconstruct historical long-term effective population size (Ne).”

**RESPONSE:** Done.

Line 50-52: Please, consider changing the focus of the definition of Ne towards historical long-term Ne as this is the parameter that is inferred in the paper.

**RESPONSE:** This is just the definition of Ne. We have included that we are inferring historic Ne. We hope that it's now clearer to the reader.

Line 53 and 56: Just a suggestion: consider giving the ranges for H and Mean Ne in the text and the mean and SD in parentheses.

**RESPONSE:** Done.

Line 53-54: Consider rephrasing to emphasise that are trajectories over time: “Trajectories of changes in Ne through time for all 68 species were estimated...”.

**RESPONSE:** Done.

Line 56-58: It would help the reader to add the purpose of the modelling at the end of the sentence.

**RESPONSE:** We rephrased the sentence to include the purpose of the modeling.

Line 62-63: The island-continental species distinction is stated here but not explored in the paper. Maybe it is not needed to state it here as it may mislead the reader towards thinking that this distinction will be relevant on the paper.

**RESPONSE:** We agree with the reviewer that this is a little confusing and have removed the information about island species.

Line 63: Reword: “long-term conservation concern” (no “concerns”).

**RESPONSE:** Done.

General: Make sure to capitalise all IUCN Red List categories, as recommended by the IUCN Red List.

**RESPONSE:** Done.

Line 68: Include Critically Endangered in the threatened category.

**RESPONSE:** Done.

Line 69-70: Add post-hoc Tukey’s Honestly Significant Difference tests p-values in parentheses.

**RESPONSE:** Done.

Line 70-71: Rephrase here as the linear models presented here use IUCN Red List category as a predictor and not the other way round as suggested here.

**RESPONSE:** Done.

Line 71: Soften the language here: “are likely because” or similar instead of “are because”.

**RESPONSE:** Done.

Line 73: Add a comma: “to mutation rate [17], changes...”

**RESPONSE:** Done.

Line 74: specify that this refers to changes in Ne among species and not changes in Ne through time.

**RESPONSE:** Done.

Line 81: I find the use of the term cryptic species a little bit confusing here. Wouldn't it be more appropriate to use something like species with the Data Deficient category as you discuss in Line 151?

**RESPONSE:** We have added to the sentence stating that this could be Data Deficient species.

Line 91: I would add a generally or something similar in this sentence: "generally resulting in population declines during...".

**RESPONSE:** Done.

Line 126: I suggest removing the "For instance," as these two sentences talk about different things.

**RESPONSE:** We have kept the "For instance" as the second sentence is elaborating on the statement in the prior sentence.

Line 132: Consider changing this sentence to something like: We measured correlation between Ne and H with body mass and diet for clarity.

**RESPONSE:** Done.

Line 139: What do you mean by genomic viability?

**RESPONSE:** Thanks for catching this typo. It should be "variation" not "viability". We have changed the word.

Line 176 and 328: I could not access the github link. I guess it is still not public. Please, make sure you make it public.

**RESPONSE:** Once the paper is accepted we will make the repository public.

Line 215-216: Consider expanding a little bit more what do you mean by: "The dataset represent the species available that passed our downstream QC criteria". I guess you mean avian species and what were the downstream QC criteria? Also change "represent" by "represents".

**RESPONSE:** We have changed the paragraph and removed the confusing sentence.

Line 235: Remove "the" from "the genome-wide heterozygosity".

**RESPONSE:** Done.

Line 240: Consider adding a brief description of how SNPs were called and filtered for PSMC analyses.

**RESPONSE:** We have added a sentence describing the SNP calling and additional filtering for PSMC.

Line 280-281: Does that mean that there may be species with as few as 31 records that have been included? Or did you mean fewer than 3000 records?

**RESPONSE:** For some species only a small number of records that fit our criteria were available. While addressing this question we found that our cutoff of 30 records had been applied before further subsetting available points to only those included within the BirdLife distribution polygons (as a further check on accuracy). Thus, we actually included species with as few as 17 points. We've corrected this in the methods. While this is a relatively small sample size, we feel it is acceptable for several reasons. First, 84% of species (57/68) had more than 50 points, and the average was 1549. So the majority had ENMs based on very large numbers of points. Second, species with small sample sizes tended to be endangered and/or be found only in small areas (e.g. the kiwis), resulting in smaller numbers of observations, but we felt it was important to include species with these characteristics in the study. Third, a similar study comparing Ne and ENM-derived habitat area (Chattopadhyay et al. 2019, "Fluctuating fortunes: genomes and habitat reconstructions reveal global climate-mediated changes in bats' genetic diversity", Proceedings B) included species in their modeling with as few as 11 records. Finally, we re-ran models excluding the species with small numbers of records (<50) and got essentially the same results. We have added some of this information to the methods.

Line 312: There is a missing word between "was" and "from".

**RESPONSE:** Done.

## FIGURES

Figure 1: Just a suggestion: I don't find drawings of individual species that fall into that category very informative. I would change the drawings by the IUCN Red List symbols.

**RESPONSE:** We appreciate the comment but have chosen to keep the bird drawings.

Figure 2: What are the drawings representing? If the species were chosen for a particular reason, it should be explained in the figure caption.

**RESPONSE:** The bird drawings are examples of bird species with suitable habitat area near the minimum and maximum observed  $H$  and  $N_e$ . We have added information to the figure texts on what the bird drawings represent.

Figure 3: I would suggest adding the normalised area as in SI figures.

**RESPONSE:** We appreciate the comment but find the normalised areas add a lot of clutter to the figures, which we want to minimize in the main body of the manuscript. The normalised area versions of these figures are available in the SI.

SI 3: Consider expanding the caption to make it more informative. Also change "normalized" to "normalised" in all plots.

**RESPONSE:** Done.

SI 5: genome and SRA codes are not shown. Please, fill in the table.

**RESPONSE:** Done.

SI 6: the column "# sites with no repeats" appears as one order of magnitude smaller than it would be expected. Please, double check the numbers.

**RESPONSE:** Thanks for catching this typo. We have changed the column name to "# sites with repeats".

Referee: 2

### Comments to the Author(s)

This is an interesting study on the power of whole-genome analysis of single individuals to advance our understanding of important demographic parameters. The authors estimate long-term effective population size ( $N_e$ ) and present heterozygosity ( $H$ ) of 68 bird species from whole-genome data, and explore the relationship between these two parameters and several natural history and conservation variables. They find a positive correlation between the area of suitable habitat and  $N_e$  (but not  $H$ ), confirm a negative correlation between body mass and  $N_e$  and  $H$ , and as expected they find that carnivores had lower  $N_e$  and  $H$  than primary consumers.

The results are not necessarily novel, and most patterns have been demonstrated before with more limited molecular datasets. For example, it is well established by phylogeographic studies that post-glacial expansions into northern latitudes from southern refugia have led to temperate and boreal populations of organisms with large effective population sizes but low genetic diversity. However, it is remarkable that whole-genome sequence data from a single individual can confirm results that until now required the use of hundreds of individuals. The PSMC results also provide information on  $N_e$  for thousands of generations into the past, which can provide interesting associations between population size and geoclimatic events. The number of species used and the taxonomic representation is also impressive.

The link between patterns of  $N_e$  and  $H$  and biodiversity conservation I find less compelling. The negative relationship between genomic  $H$  and extinction risk is interesting, yet the resolution is not high, and critically endangered species are not significantly different from vulnerable species. So the suggestion that genome-wide estimates of  $H$  can be used as indicators of conservation status seems premature and overly optimistic. Surely conservation status can be done more easily (less expensively) with conventional genetic markers? In any case, I would tone down the direct relevance of these results to conservation,

perhaps not referring to it in the title.

**RESPONSE:** We agree with the reviewer that traditional approaches (i.e., microsatellites) are informative and less expensive, but as the field of genomics is rapidly advancing and the cost associated with NGS sequencing dropping, we do think that investigating how these data can be used to inform conservation is important. That the resolution between the threatened Red List categories (e.g., Vulnerable, Endangered, Critically Endangered) are low is not surprising. The 68 species included in the manuscript showed large variation in  $H$  and we would expect that increasing the sample size would add higher resolution. The interesting finding here is that we do in fact identify overall declines in  $H$  for threatened and non-threatened species.

Regarding the long-term  $N_e$  estimates from PSMC (Figure 3), error in recent periods seems too high for the last 10000 to 100000 years to provide reliable data that can be associated with recent climatic events, at least for species with long generation times. Perhaps the Holocene should be excluded as little can be said from these analyses for such a recent period. Also, the two species selected for this figure are eminently tropical, yet the effects of the historical events used (Holocene, Last Glacial Max, Last Interglacial) had the most severe effects on temperate and boreal species. Perhaps including a temperate/boreal example would be more informative.

**RESPONSE:** We agree that PSMC error is high for some species in recent years. As noted in the methods we removed recent PSMC estimates when calculating the mean  $N_e$  value used in our modeling. However we have decided to keep these values visible in the PSMC plots (Figure 3 and SI 3) so readers have all the PSMC information available to them. We replaced one of the species in Figure 3 with a temperate species, American crow.

For results in Figures 2 and 4, please include in the legend or in the main text the regression R squared values for all plots.

**RESPONSE:** As we note in the figure captions, the information in figures 2 and 4 comes from multiple regression models (one for  $H$  and one for  $N_e$ ). R squared values are available for each model in total, which we have now provided in the text. It is unclear how valid partial  $R^2$  approaches (that is, calculating  $R^2$  on a per-covariate level from a multiple regression model) would be for our updated models fit with PGLS on request of the other reviewer. Thus we have not included partial  $R^2$  values on each figure.

## Appendix B

Dear Dr. Costa,

Thank you for sending the reviewer comments on our manuscript *Life history traits and habitat availability shape genomic diversity in birds: implications for conservation*. We have carefully revised the manuscript according to the reviewer suggestions. We hope that this revised version is now acceptable for publication in Proceedings of the Royal Society B.

Best,  
Anna Brüniche-Olsen

Reviewer(s)' Comments to Author:

Referee: 2

Comments to the Author(s).

I appreciate the changes made to the previous version, which I think have improved the manuscript. However, regarding Figure 3, which combines the PSMC plots with maps of suitable habitat (ENM), I think they remain misleading. The  $N_e$  estimates for the Holocene (vertical green bar on Fig 3) are not reliable and typically wrong. This is clear in the new figure for the American crow. There is a clear expansion of habitat area in the Holocene following the Last Glacial Maximum, yet the subsequent increase in  $N_e$  is not detected by PSMC. I think the authors have two options, either remove the Holocene maps from these plots, or show the increase in the variance of  $N_e$  estimates at these recent dates by showing the bootstrap replicates for each estimate, which have been cut on every plot for the last 10000 years (without justification or explanation). As the plot appears right now, the reader must interpret that following the Last Glacial Maximum, there has been an increase in suitable habitat but not an increase in  $N_e$ , which is not true. Adding the bootstrap replicate lines to the entire plot would at least alert the reader of the large error associated with recent estimates of  $N_e$ .

**RESPONSE:** We thank the reviewer for identifying this issue. We have retained the Holocene maps in the plots, and have fixed a mistake in our plotting code which was hiding the bootstrap lines for the Holocene period. All figures now include complete bootstrap lines. In some cases they are hard to see because they overlap the overall PSMC estimate so closely (this is true for both species in figure 3), but for other species in the appendix plots the bootstrap replicates are very wide as the reviewer suggested they would be.

There is a sentence about the uncertainty in recent  $N_e$  estimates in the manuscript describing that we exclude these from the model. Sentence reads (L335-337) “Very recent estimates of  $N_e$  from PSMC tend to be noisy thus following Leroy et al. [73], we excluded the four most recent timepoints of  $N_e$  for each species when calculating the mean estimate to increase reliability.”

Referee: 3

Comments to the Author(s).

In their manuscript, Brüniche-Olsen and colleagues combine whole genome and ecological data to understand the role of biological and environmental factors in determining levels of genetic diversity to inform conservation. I find that their approach is original, and the paper is well written and accessible to a broad readership. I particularly appreciate the use of some of the genomes available as it shows how much we can infer about a species from the genome sequence of a single individual. As this is the first time I review this manuscript, I think the authors did a good job addressing previous comments from two reviewers. My major complaint is that this version of the manuscript didn't include line numbers. Please find below some minor comments/suggestions.

**RESPONSE:** We are sorry that we forgot to include line numbers in the uploaded version. We agree that this makes it harder for reviewers to indicate where they have comments to the manuscript. We apologize and appreciate that the reviewer took time to provide comments regardless.

End of page 4: could you repeat what the covariates are?

**RESPONSE:** We have added the covariates to the sentence. It now reads (L79-81) "Our aim was that the results would identify how different extrinsic (suitable habitat) and intrinsic (diet, body size) factors impact genomic diversity, and may ultimately be used as reference to inform conservation."

Page 6: generally "cryptic species" has a different meaning as it refers to species that haven't been identified as distinct from another one.

**RESPONSE:** We have rephrased the sentence and refer to Data Deficient species. It now reads (L127-129) "...insights into species listed as Data Deficient by IUCN. For instance, we found that 66% of species with  $H < 9 \times 10^{-4}$  were categorized by IUCN as Vulnerable or Endangered; thus, a Data Deficient species with..."

Page 7: this reference to a 'newly discovered baleen whale' is not clear. It reads like the baleen whale has been recently discovered. I suggest replacing with 'a previously undescribed baleen whale species (Balaenoptera spp.)'

**RESPONSE:** We have changed accordingly.

Page 8: 'The myriad species' is surely an overstatement when referring to 68 species.

**RESPONSE:** We agree that this is an overstatement and have removed "myriad". The sentence now reads: (L153) "The species studied.....".

Page 8: I wouldn't consider diet a life history trait per se, but perhaps a proxy for a suite of life history traits. Additionally, aren't body mass and diet highly correlated?

**RESPONSE:** For the sake of simplicity we have put diet in the "life history category" in this paper. There was some correlation between body mass and diet in our dataset (see figure below). However there was a high degree of variability in body mass within each diet type. Thus, we felt it was appropriate to consider them separately in our modeling.

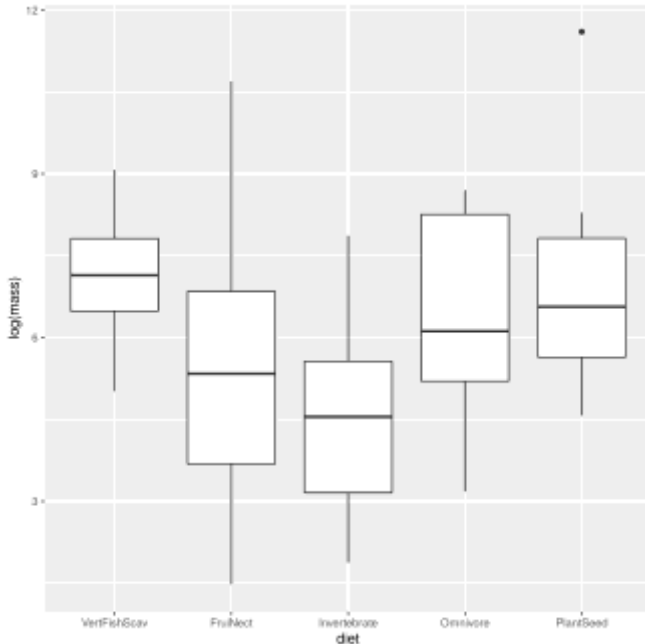


Fig.2 Missing arrow associated with species on the left in the left side panel. Also, please add species depicted in plots in the figure caption.

**RESPONSE:** Thanks for catching the missing arrow. It has now been added. We have also updated the figure text to include the names of the species shown in the plot and in all other figures.

Page 16: why was ANGSD used for estimating heterozygosity and how? The depth of the data used here should be high for all these species and the SFS for a single individual doesn't really help with estimating heterozygosity in this case. Was heterozygosity the sum of variant sites divided by genome size or by the actual proportion of the genome covered by high quality data? Did you apply some high and low coverage filters?

**RESPONSE:** We used ANGSD as it work for both high and low sequencing depth. We agree that we could have called genotypes and analyzed those, but ANGSD is equally good so for this it's a matter of choice.

We have extended the description of how ANGSD was used for heterozygosity estimation. The section now reads (L289-294) "For each species, we quantified genome-wide heterozygosity (H) based on the site frequency spectrum (SFS) using ANGSD [35]. This H measures the proportion of heterozygous genotypes divided by the genome size while excluding sites of low quality. We used filters on base quality score (-minQ 20), minimum mapping quality (-minMapQ 30), and mapping depth (d) setting minimum mapping depth to 1/3 of the mean total mapping depth (-setMinDepth d/3), and maximum mapping to double (-setMaxDepth d\*2) the mean total mapping depth."

## Appendix C

Dear Editor,

Thank you for the reviewers comment. We have addressed the comment and hope that our manuscript is now ready for acceptance in Proceedings of the Royal Society B.

Sincerely,  
Anna Brüniche-Olsen

Associate Editor

Board Member: 1

Comments to Author:

I would like to say thanks to the authors for their efforts in improving the paper. Most of the previous concerns have been addressed. However, the issue raised by the referee about the impact of removing effective population size datapoints for the Holocene is significant and should be addressed before publication.

**RESPONSE:** We have clarified this point below.

Reviewer(s)' Comments to Author:

Referee: 2

Comments to the Author(s)

I appreciate the authors' reply to my comment, but I am afraid their arguments are insufficient and the main problem remains. They explain in the methods that "Very recent estimates of  $N_e$  from PSMC tend to be noisy thus following Leroy et al. [73], we excluded the four most recent time points of  $N_e$  for each species when calculating the mean estimate to increase reliability." However, Leroy et al. (2021) removed those four recent points because they were interested in obtaining an average  $N_e$  value for the last million years, and were not specifically interested in the most recent few thousand years, which is where PSMC is unreliable. But here the situation is clearly different, as the PSMC is used specifically to infer  $N_e$  during the Holocene, and establishing a direct comparison with the extent of habitat availability that is actually depicted in a map within the same figure. If the authors remove the four most recent time intervals, then there is actually no data for the Holocene, and that is why the plot line is flat on the left end on all species. In my last review I requested that bootstrap lines be added to all time intervals to reveal variance, but it is now clear that the data were removed, so the problem is more serious (and the bootstrap lines are meaningless). Ideally the dataset would include several genomes per species, instead of just one, so that other approaches could be used to increase the precision of  $N_e$  estimates in recent time periods (msmc, momi2, etc.). But given that only one genome is available per species, I think the plots should either show the  $N_e$  values for the recent time intervals (revealing noisy variation), or exclude the Holocene maps in the lower panels, as no reliable data seem to be available for that period.

**RESPONSE:** Reading the reviewers comment we realize that we needed to further clarify how we have approached the comparison of  $N_e$  and habitat availability in the multiple regression analysis, and what is included in the PSMC trajectory plots. As mentioned, we used the same



approach as Leroy et al (2021) by estimating an average  $N_e$  across the past 1 million years (excluding the four most recent points). This was done as there are inherent uncertainties in PSMC in regards to parameterization choices (e.g., generation times and substitution rates) which may vertically shift the PSMC trajectory. We wished to limit these biases as much as possible and therefore we did not focus on  $N_e$  for single time periods but instead used an average  $N_e$  and an average of the suitable habitat over time in our regression analysis. The goal here was to reduce the influence of those uncertain recent points, which is line with what Leroy et al (2021) did. To summarize, in the regression analysis (the results of which are in Figures 1, 2, and 4) we did not consider anything specifically about the Holocene or any of the other time periods; we considered the entire time periods (after removing the four most recent points) for both  $N_e$  and habitat availability. We have clarified how the  $N_e$  estimates were used in L323-329 and in SI 2 table text.

However, in the PSMC trajectory plots (e.g. in Figure 3 and SI 3), we did *not* remove the four most recent data points. We emphasize this in line 327-329. These figures retain all the available data, and all available bootstraps; we have now clarified that the bootstraps are included in the plots in L238-246. As the reviewer notes, some of the most recent data points on some of the PSMC plots include few coalescent events and thus produce flat lines throughout the Holocene. In other cases, the uncertainties associated with shallow time were reflected in wide variation among the bootstrap values. We think it is best to illustrate these uncertain recent data points in *all* the PSMC plots so that readers can assess the data for individual species themselves. We have provided some further clarification and cautions on how to interpret these PSMC lines in the captions for SI 3 and Figure 3 (L 245-248).

We agree with the reviewer that having multiple genome sequences per species would be ideal, and it will hopefully be feasible as more and more wildlife species gets sequenced. In our study we were limited by the amount of genome data available at public databases and therefore used a single genome per species. Despite this limitation, we think our study effectively demonstrates that a) avian life history traits shape their contemporary genomic diversity and b) mean heterozygosity serves as a leading indicator of demographic change and thus has great potential for use in conservation monitoring. We think the readers of PRS will appreciate these big-picture findings and hope you agree.