9th April 2025

ROYAL (DICK) SCHOOL OF VETERINARY STUDIES
The University of Edinburgh
Easter Bush Campus
Midlothian
EH25 9RG
UNITED KINGDOM

Dear Professor Nabholz,

Please find enclosed the resubmission of our manuscript entitled '**Population structure and genetic diversity of the Critically Endangered bowmouth guitarfish (*Rhina ancylostomus*) in the Northerwest Indian Ocean'**. We are very grateful to yourself and the three reviewers for the helpful and constructive comments and include below details of how we have dealt with each one. The comments are in blue, and our response is in black.

Please let us know if you require any further information.

Yours sincerely,

Dr Emily Humble and colleagues

Editor Comments to Author:

Your manuscript was evaluated by three reviewers. Reviewer #1 and #3 are positive and all appreciated your work. I largely agree and believe that the use of a powerful genomic method to better understand the population structure of an endangered shark species is very valuable and of great importance for conservation.

Reviewer #1 offers good advice for improving your text. I noticed that you have already addressed the comment regarding the low genetic diversity of the CR region in the discussion section.

Reviewer #2 provides numerous comments and focuses on sampling issues. The limitation of not sampling the Indo-Malayan archipelago could be discussed further. For instance, you could expand on references to coastal reef shark studies if they support the reviewer's suggestion that the Indo-Malayan regions often harbor greater genetic diversity. Additionally, the imbalance in sampling between localities is noted. I suggest performing a control analysis by randomly subsampling all populations to N=4 or 5. This would allow you to test the effect of unbalanced sampling on your conclusions. Indeed, unbalanced sampling has been shown to influence structure-like analyses (Puechmaille 2016) and PCA. Finally, several comments are made regarding your discussion. While I do not fully agree with all of them, I concur that you could directly interpret your results as supporting a clinal variation from east to west with discontinuous sampling creating gaps. Therefore, I suggest modifying your text accordingly in both the results and discussion sections (e.g., lines 337-341).

Reviewer #3 is positive. He wants clarification on the potential ascertainment or allelic bias of the nuclear marker. He also adds a comment on the clinal nature of the genetic variation, which supports the modifications suggested above.

Please respond to all the comments made by the reviewers, but if you choose not to accept some of them, please justify your decision in the replies.

As a final note, I haven’t found the supplementary figures and tables to be available. Could you please make them available in the next version of your preprint?

Reference
Puechmaille SJ. 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Mol. Ecol. Resour. 16:608–627.

We thank the editor for their encouraging and constructive comments on our manuscript. We have now responded to all suggestions made by the reviewers as well as those summarised by yourself. An updated version of our Supplementary Material has been uploaded on bioRxiv and to this submission. Overall, we believe our manuscript is stronger because of the changes and hope it will be accepted in PCIEvolBiol.

Review by anonymous reviewer 1, 29 May 2024 14:13

The manuscript from Kipperman et al. investigates the population structure of the bowmouth guitarfish at both mitochondrial (control region and COI) and nuclear (DART sequencing) level. The bowmouth guitarfish represents an interesting case study due to its status (Critically Endangered in the IUCN Red list of Threatened species) and its geographic distribution.

I do not see any major flows in the data generation process and population genetics analyses. However, the sampling strategy is not optimal. I do understand that it is very difficult to perform an adequate and extensive sampling for marine species and I can imagine that adding more individuals

would be impossible at this stage, but the data at hand are unfortunately limited. As consequence, conclusions do not seem highly robust and are a bit limited in scope. There are three problems for me concerning the sampling: the low number of individuals sequenced, the unbalanced sampling (67% of the whole nuclear samples comes from the same locality, UAE), the geographic locations not covering much of the total distribution and typically ignoring regions which are most likely harboring largest diversity (the Indo Malayan archipelago) following many biogeographic studies and the genetic diversity in coastal reef sharks.

We acknowledge the reviewer's concern about sampling limitations but as they point out, adding additional individuals and regions to the study would be extremely challenging at this stage. Our sampling strategy was deliberately comprehensive within logistical constraints – we aimed to have representation from as many accessible regions where this rare and endangered species occurs. While we recognise that further sampling remains a priority for future research, the pressing need for evidence-based management decisions for this vulnerable species outweighs the marginal benefits of waiting for what could be only a few additional samples. To address the reviewer's concerns, we have updated several sections in the discussion to further highlight our sampling design limitations (L380 / L403 / L440) while contextualizing these challenges within the reality of research on endangered marine species. Additionally, we have conducted supplementary analysis on a balanced subset of samples to further validate our findings (detailed in our response below).

With regard to the Indo-Malayan archipelago, we strongly agree with the reviewer about its significance for the species. It was by no means a location we 'ignored' but rather somewhere that was especially challenging to procure samples from within the timeframe of our research project. The final sentence of our manuscript explicitly emphasizes the importance of extending this research to include this critical region.

I have some remarks about the discussion, which sometimes is unnecessary long and disconnected from evidence presented in this work.

Lines 320-321. It is unclear what do they mean with 'contemporary evolutionary constraints'. Also, past directional selection should affect also the COI since the mtDNA is a single linkage block. Also it is not clear what biases in mutation type are and how should they contribute to determine the observed pattern.

Low control region variation is a pattern that has been observed in several elasmobranchs and several teleosts. It has been hypothesised that functional constraints for a particular base composition in the control region has led to a lower probability of transition mutations and therefore lower variation overall (Apostolidis et al., 1997; Bernatchez & Danzmann, 1993). We have now clarified this explanation in the Discussion (L366).

Lines 328-330. First, the observed pattern may partially be driven from the unbalanced sampling. Second, it seems from their results that there is a clinal distribution of the genetic diversity, implying continuous gene flow driven by geographic distances. At best, this sentence should be reformulated by stressing the evidence of continuous variation in diversity.

We have now reanalysed a balanced subset of our data which has revealed no difference to our overall results. We describe this process in our Materials and Methods (L206–212) and present the findings in the Supplementary Material (Figures S7–S9).

In response to the reviewer's second point, we fully agree and have slightly reworded this sentence. We would also like to point out that in the second half of the paragraph, we explicitly state how we expect the species to display continuous variation in genetic diversity: "We therefore cautiously anticipate that the species displays a pattern of isolation by distance across much of the range assessed here with clinal variation in allele frequencies being driven by intrinsic dispersal

constraints." This interpretation is also highlighted in the final section of the discussion and the abstract.

Lines 330-332. I do not understand the meaning of this sentence. The analyses do not suggest the presence of two differentiated clusters but, once again, of a clinal variation. Also, Fst analyses are not extremely useful in this context given the unbalanced sampling and the very low number of individuals of some locations. I suggest removing the Fst because of this sampling scheme.

In this sentence, we are simply highlighting how our SNP-based structure analysis, PCA and Fst uncover observable differentiation between locations. This is quickly followed by a caveat highlighting how these clusters are likely to be artificial due to gaps in our sampling distribution. We have rewritten this section for clarity (L393–399). With regard to low sample numbers, rather than removing the analysis which we feel contains useful information, we further emphasise the limitations of low sample numbers in our Results (L311–312) and Discussion (L380 / L403 / L440).

Lines 337-339. I do not understand this sentence. The authors observed linear differentiation, which is a form of population structure. I do not see the antithesis with "true population differentiation". Please remove of reframe this sentence.

We have rewritten this sentence. It now reads: "The population structure observed in our dataset is therefore most likely driven by gaps in our sampling distribution as opposed to true population boundaries."

Lines 352-363. Not sure this paragraph adds much to the manuscript

We feel strongly that this section should remain in the manuscript. Although it is quite speculative, it highlights some key scenarios that will be important to consider when designing both *in situ* and *ex situ* management plans.

Line 370. The species is present only in the IP, so clearly it would be impossible to infer its origin anywhere else. But I agree that given the sampling, it is impossible to infer its centre of origin correctly

Thanks for pointing this out, we meant the Indo-Malayan archipelago and have corrected our text accordingly.

Lines 387-388. It seems that gene flow is ongoing but this has not been formally tested (populations can be recently become isolated, for example). Not sure this sentence adds much.

This is an important interpretation of our findings and we feel strongly that it should remain.

Lines 390-399. This applies to any species. It is unclear why the results presented here should prompt more conservation effort than in any other species. Maintaining habitat connectivity seems to be always crucial, but this is quite unlinked from the results presented here.

First, we are not advocating for more conservation effort than in any other species but rather a suite of approaches that might be appropriate for *Rhina ancylostomus* given our genetic findings. We have clarified this in the text.

Second, maintaining habitat connectivity is not necessarily always a priority. For example, some marine species display very strong population structure with restricted gene flow across very large distributions. In these cases, maintaining habitat connectivity across the entire species range would not be beneficial. We show how *R. ancylostomus* most likely displays a pattern of isolation by distance across its northern range, where gene flow occurs but is limited by the constraints of dispersal. We feel this provides a strong argument for large-scale habitat connectivity, particularly for the maintenance of genetic variation.

Minor comment:

Line 180: it is unclear how the bootstrap is performed. Is it the same permutation approach of Excoffier et al. 1992?

Pairwise F_{ST} values and bootstraps were calculated according to the method proposed by Wright (1949) and updated by Weir and Cockerham (1984). We have clarified this in the text.

Review by anonymous reviewer 2, 14 May 2024 21:43

Introduction

Line 51-54: You started the paragraph mentioning the high level of gene flow in large marine organisms and then the given example is talking about inbreeding – which is very unlikely to happen when there is high gene flow occurring. I suggest modifying this paragraph by potentially including the first two sentences of paragraph two together with the last sentence of paragraph one and having it as a separate paragraph.

Thanks for raising this point. We have rewritten the second paragraph of our introduction to address this.

Line 57: "fantastic opportunity to explore these problems" – What problems? You need to make it clear to your reader.

Thanks for pointing this out. This sentence now reads: "Genetic and genomic tools provide a fantastic opportunity to explore the landscape of genetic diversity and differentiation and are increasingly being applied to elusive marine megafauna."

Methods

Line 136 – Explain in more details how and why those enzymes were chosen for the digestion.

The enzymes were chosen by DArT for their ability to isolate highly informative, low copy fragments of the genome in a reproducible manner. We have now included this information the methods.

Line 185-197 "We determined geographic distances based on a least-cost path analysis using the R package marmap (Pante & Simon-Bouhet, 2013)" – Since the sampling description of your samples say that they were obtained from fisheries, I would like to know what type of geographic information was used to calculate the "geographic distance". Opportunistic sampling usually does not offer geographic information. Please, explain this in your methods a little bit more.

We thank the reviewer for bringing this up as it is an important point. In this case, we used the geographic location of the coastline directly adjacent to the landing site in which the samples were collected. This is of course not a true approximation of where the individual was caught. However, *Rhina ancylostomus* is a coastal species and therefore we thought it reasonable to assume this would be a good approximation of their geographic location. Given that we uncover geographic signal in our dataset across multiple methods, we believe this is a fair assumption to have made. We have clarified the origin of geographic information in the text (L215–219).

Results

Table 1 – Interesting to see that the COI presented a lot more number of haplotypes and diversity than the CR region.

This is correct and in response to reviewer comments we have now elaborated on this result in our discussion.

Line 239-241 "For the CR, our sample size of 65 individuals was sufficient to recover both 95% and 99% of the haplotype diversity in the species (Figure S2C–D)". Would you have an idea why that might be considering that the CR is usually the most polymorphic region in the mDNA? Perhaps the

CR region where the primer was designed do not cover the polymorphic region of this gene? This should definitely be considered when designing a primer.

This is a pattern that has now been observed in several elasmobranchs and many teleosts. It has been hypothesised that low CR variation could be driven due to differences in the ratio transition/transversions, underpinned by functional constraints for a particular base composition (Apostolidis et al., 1997; Bernatchez & Danzmann, 1993). We have now expanded on this explanation in the discussion although do not go into too much detail since it is beyond the scope of our study. This aside, it is also possible we invertedly targeted a conserved region of the CR during primer design which we also highlight in the discussion.

Line 247-250 – For better visualization, I suggest adding the pairwise Fst figure for the CR region in Figure 2. CR is the most used mtDNA marker in population genetic studies, so I think displaying it as the main result is necessary.

We understand the reviewers point however in our study the control region presented little variation across our samples and therefore was not suitable for visualisation of Fst.

Line 265-267 – I do not think is necessary to display the figures referring to K>2. I would suggest displaying only the K=2 as it is the optimal value for your samples.

We have updated the figure to only include the individual assignment plot for K=2. We have moved the remaining figures into the Supplementary Materials.

Table 4 – UAE and Oman seem to have low genetic diversity considering the number of samples analyzed for these locations compared to Bangladesh and Saudi Arabia.

This is true for the control region but not for COI, where Saudi Arabia displayed no variation. When combining these results with those from our SNP data and while accounting for the variation in sample size, the most consistent finding was a decline in variation from east to west of the sampling range. We cautiously interpret this as a signal of possible range expansion.

Discussion

Line 312-325 – What was taken into consideration when targeting the CR region during primer design? The CR region has approximately >1000bp and most of the population genetic studies targeting elasmobranchs cover at least 750bp. I suggest modifying the discussion on this topic to account for the technical limitation of this study. You brought up several reasons of why this might be happening but did not provide references to support your statement (Lines 319-321).

As stated above, the reviewer is correct that we may have invertedly targeted a conserved region of the control region during primer design. We have included this caveat in the discussion as well as highlighted the value of complete gene regions and whole mitogenomes for both population genomic analysis and for investigating mitochondrial evolutionary rate variation. We have also provided more detail and references around the potential drivers for low sequence variation.

Line 333-337 – Very good observation.

Line 372 – 375 - It is not ideal to make this type of affirmations or suggestions when the study do not present a standardized sample size. For the SNPs, Saudi Arabia is only represented by two samples. I suggest exploring more this reference Domingues et al. 2018 for this paragraph.

We agree that having only two samples from Saudi Arabia presents a limitation however we feel they are valuable data points representing an important region. Furthermore, the cline in genetic diversity remains when Saudi Arabia is removed from the picture. Nevertheless, we have toned down the statement referred to by the reviewer and have included further recognition of the need for

additional sampling in order to validate the findings presented. We have also addressed the issue of unbalanced sampling by reanalysing a subset of our data (see above)

Review by anonymous reviewer 3, 02 Jul 2024 05:49

The authors present a conservation genetics study of the bowmouth guitarfish, in which they use mitochondrial markers and SNPs from a nuclear SNP-typing array to assess population structure over much, but not all, of the species range.

There are not many details on the SNPtyping, for example how SNPs were ascertained, and how/whether the authors test for ascertainment or allelic bias. It would be useful to have such details to be order to assess the structure results from nuclear markers.

SNPs were genotyped by DArT using DArTsoft, a proprietary bioinformatic pipeline developed specifically for DArTseq data. We are therefore unable to describe this process in the same level of detail as other parts of our Materials and Methods. This is stated in our manuscript, and we reference the original paper describing the DarTsoft pipeline.

All subsequent SNP filtering was carried out by us and described and justified in detail in our methods section. Furthermore, all the analysis code required to reproduce each step of our SNP filtering pipeline (and subsequent analysis) is provided in a documented GitHub repository.

Sample sizes are uneven among putative demes, and often small, which the authors acknowledge and address.

We have now addressed this further by presenting a reanalysis of a more balanced dataset (see response to Reviewer 1).

The pattern of weak structuring reflects recent findings in some dolphin species (Gose et al. 2024) by a team that includes the lead author of this study, and perhaps this is more the norm in this high connectivity environment. I did wonder whether the apparent break point in the structuring between SL and UAE is due to the lack of sampling and geographic distance between the two, so that the reality is this even more of a cline than suggested, rather than a case of clear structuring.

Yes exactly. We argue this in our discussion and have now further emphasised the limitations to our sampling constraints.

Overall, the authors have performed a suite of analyses that complement each other, and have judiciously assessed the strengths and weaknesses of the resulting findings of each. This study is clearly of high importance for the conservation efforts to manage this critically endangered species.

I therefore recommend the study for publication.

Thank you very much for your positive assessment of our manuscript.

References

Apostolidis, A. P., Triantaphyllidis, C., Kouvatsi, A., & Economidis, P. S. (1997). Mitochondrial DNA sequence variation and phylogeography among Salmo trutta L. (Greek brown trout) populations. *Molecular Ecology*, *6*(6), 531–542. https://doi.org/10.1046/j.1365-294X.1997.d01-176.x

Bernatchez, L., & Danzmann, R. (1993). Congruence in control-region sequence and restriction-site variation in mitochondrial DNA of Brook Charr (Salvelinus Fontinalis Mitchill). *Molecular Biology and Evolution*, *10*(5), 1002. https://doi.org/10.1093/oxfordjournals.molbev.a040062