

## Reviewer and editor responses

Dear authors,

The manuscript has now gone through a second round of revision by the same three reviewers as the first round. All the reviewers have stated again the great merits of this study, and I share this point of view. All also felt a bit disappointed that the authors did not try (or were not able to try) an alternative method to hPSMC that can handle both complete lineage sorting AND gene flow at the same time and test how comparable results are. Reviewer #1 (Christelle Fraisse, CF) suggested to try MSMC-IM even on a subset of the species for which suitable data would be available, and test how comparable would be the results from hPSMC and this methods. If this is not possible, may I suggest to try the new Approximate IM model of Müller et al. (Joint inference of species histories and gene flow. bioRxiv 348391; doi: <https://doi.org/10.1101/348391>) which can handle one single reference genome per species and infer jointly the species tree, the split time among species, and gene flow among them while accounting for ILS and gene flow. Either way could really strengthen the results, the scope, and the conclusions of this study and address at least in part the reserves raised by reviewer #1 (CF) and reviewer 3 (by Andy Foote).

Beside that, some minor issues were pointed by reviewer #1 (CF) that can be fixed easily.

Assuming the authors would be able to validate the results of hPSMC with an independent approach, this would become a very strong study that I would be proud and honored to recommend.

I am staying at the authors disposal if they have any further questions and comments.

All the best

Michael C. FONTAINE

**Answer.** We would like to thank Michael for his positive comments and guidance during the review process. We have now added an additional analysis using relative divergence times of alternative topologies within the Delphinidae to assess the robustness of our results, and due to the additional expertise requirements, our revised manuscript therefore also includes co-author David Duchene. The results from this analysis were highly convergent with those of the hPSMC results suggesting that the speciation process within this family was a long process. Unfortunately as it is very difficult to fully disentangle ILS from gene flow, we were unable to confidently comment on which phenomenon was causing the discordances. We have therefore changed our wording to be more cautious but also added some additional information as to why it seems more likely that gene flow explains our results as opposed to simply ILS. Although we were unable to implement all the suggested analyses, we hope that this new analysis now adds to the robustness of our previous results and makes our manuscript suitable for your recommendation.

## Reviews

Reviewed by Christelle Fraïsse, 2021-04-08 14:48

The authors made efforts to address most of the concerns raised during the reviewing process, which is a positive point. Especially, they:

- i) added a new tree topology-based approach (QuIBL).
- ii) tested the robustness of the tree topology-based approach to GC content.
- iii) tested the robustness of the hPSMC approach to repeat regions.
- iv) they did bootstraps of the hPSMC for some species pairs.

Despite these great additions, I was left a bit disappointed that the authors did not compare their hPSMC results with that of MSMC-IM. It would have been a way to validate the timings for the cessation of gene flow with a second method. Given the various concerns that all reviewers raised with the hPSMC method, and because the occurrence of post-divergence gene flow long after initial divergence is a crucial point made by the authors, I think it is important to consolidate this part of the manuscript. The authors justified not using MSMC-IM because they “do not have access to phased data and/or population-level data for all of our species to be able to implement these analyses in a meaningful way so as to be able to make comparisons within the superfamily as is the focus of the study.” In my opinion, testing the MSMC-IM method on a subset of species will already be informative to assess if the timings of cessation of gene flow are congruent between methods.

Apart from this point, I am very satisfied with the revisions made.

**Answer:** We understand that it would be good to have an independent source of information to support our hPSMC results. Although we were not able to implement MSMC-IM with our dataset, we have implemented an additional analysis to date when the discordant tree topologies within Delphinidae may have arisen (most likely due to ILS/gene flow). The results from this analysis are convergent with those from the hPSMC giving us more confidence in our results. Furthermore, we have also changed to more cautious wording in regards to interpreting our results as only arising due to gene flow but given more reasoning why we think gene flow was the more likely cause for delayed divergence of species within our dataset.

Minor comments on the figures:

● Figure 1: some of the cartoons are shifted, and so they do not correspond to the correct branch. Also, the grey color gradient is hard to visualize. Why not simply using the same color code as in Figure 2A.

**Answer:** We have now modified the figure by aligning the cartoons better and changed the colour coding to the same as in Figure 2A

● Figure 2: please add “divergence time” to the right of the corresponding symbol in the legend.

**Answer:** Divergence time is now found next to the corresponding symbol

Reviewed by Camille Roux, 2021-05-05 16:28

First and foremost, I apologize to the authors for my slowness in turning in the two reviews for each of the two rounds. My general opinion is that the article presented here has several interests. First, it focuses on gene flow over a broad evolutionary scale in a group of marine mammals. Allowing in the future to shed light on possible elements involved in the marine-speciation paradox by comparing with terrestrial mammals.

A second interest is that it can serve as a methodological textbook for this type of data in clades where population data are not available. A third interest is that it raises a demographic curiosity with such long periods of genetic exchange in time that are still to be understood. Of course, I'm still not fully convinced by the title. Speciation "in the face of gene flow" suggests that barriers between species have evolved in spite of gene flow. This is different from having gene flow in spite of pre-established barriers. But the authors have refined this message in the text, which is essential. In the same way, the authors considered some suggestions made, including checking for possible biases related to GC content. This is an effort on their part that should be highlighted.

I am still struck by some of the quantified results, especially the strong support of gene flow, within families, over several million years. On a personal note, I am really looking forward to seeing results using other inferential methods when data in multiple individuals will be available.

Thus, after this second round of revision, I have no further comments to make and can only support the current manuscript.

Camille Roux

**Answer:** We have now added an additional analysis to date the relative divergence dates within Delphinidae using the consensus and alternative topologies in an attempt to assess the robustness of our hPSMC results. We find that they are very concordant and believe this further adds support to the exchange of genetic material for a long time after initial divergences. Moreover, as we have lessened our emphasis on gene flow, we have also changed the title of the manuscript.

Reviewed by Andrew Foote , 2021-04-06 15:51

The authors have made some updates to their manuscript in response to the previous reviews and have added an additional analyses, based on measuring branch length variation along the genome. The conclusion drawn, that 'Speciation in the face of gene flow within the toothed whale superfamily Delphinoidea' still largely hangs on the author's interpretation of the hPSMC method.

I don't share the author's confidence that this method differentiates between gene-flow and lineage sorting. Essentially the upsweep in estimated  $N_e$  in the pseudo-diploid PSMC plots, which the authors infer as a cessation of gene flow, reflects a change in coalescence. And the inference that gene-flow continued after speciation was initiated is drawn from the comparison of the timing of the change in the trajectory of the hPSMC plot compared with the split time estimated by McGowen et al. The mapping of short read data to the distantly related baiji may be favouring the mapping to more conserved regions and may be a factor in the disagreement between the divergence estimates by McGowen et al. and the coarse estimate of the timing changes in coalescence interpreted from the PSMC plots.

I would recommend the authors to be more conservative in their interpretation of the PSMC plots of unphased pseudo-diploid genome generated by mapping short-reads to a distant relative as gene flow events that occurred millions of years ago.

**Answer:** We have now changed the wording in the hPSMC section to include the possibility that the results arose due to both ILS and gene flow. We have also commented on why we think gene flow to be the more likely reason for the delay in coalescence in the genomes of the individuals studied here. Finally we also included an additional analysis using relative divergence dates to further support the timing for the end of lineage sorting/gene flow that are convergent with the hPSMC results, giving us further support in these results. We hope that with these adjustments we have now met the reviewers' concerns.