

Dear authors,

Thank very much for having revised and resubmitted your manuscript. I have conducted a careful reading of this new version of your manuscript and I do still have a few questions, issues and points that need to be addressed before I can accept to recommend your manuscript. I have incorporated my comments and suggestions directly as comments into the pdf.

I hope this is ok with you.

I am staying at your disposal if you have any questions.

I am really looking forward reading your revised version and have the pleasure to recommend it.

Best wishes

Michael C FONTAINE

Thank you for your feedback. We have taken the comments into consideration and you will find a point by point response to each below.

**31:** This is highly speculative, I suggest replacing by "may suggest that..."

**Answer:** Changed to "suggests that" to avoid too many "may" in the same sentence.

**110:** 1- I personally find that the results/discussion section may benefit from a bit more discussion and comparison with the literature, which should not limit to only the cetacean species, since this topic is still at its infancy.

**Answer:** The aim of this study was solely to gain an understanding of the speciation process within Delphinoidea. The inclusion of a discussion of other species is outside our scope.

2- Consider making the subtitles in the results/discussion more informative on what is shown or found. Currently, they are too sparse and limited in term of information they provide to the reader.

**Answer:** We have now added more informative subtitles in the results/discussion section

**178:** Malinsky and coll. (2018, 2021) showed quite convincingly based on simulations and concrete examples from the cichlids that historical introgression among ancestors could lead to significant signal of introgression in daughter lineages, and correlated D-statistic values among related taxa, even if they are not hybridizing anymore. As such, I do not think they can be considered as "false positive" as mentioned on line 185. The evidence of excessive shared polymorphism among these daughter lineages are due to remanent of the introgression in their ancestors. The author could rephrase and precise their description and interpretations of the D-stats to be more accurate. Do not hesitate to also discuss other similar examples from other system as well as a comparison, even if these are not cetacean species.

-- Ref:

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour*, 21(2), 584-595.

doi:10.1111/1755-0998.13265

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, 2(12), 1940-1955.  
doi:10.1038/s41559-018-0717-x

**Answer:** We used the phrasing “false positive” for the exact reason you mention. If the related taxa show significant values of introgression due to introgression among ancestors, despite not actually hybridising themselves, the significance value is technically not correct. Discussions of non-cetacean species are outside the scope of this manuscript.

**184:** Are these actual limitations of the D-statistics, or instead the way they have been implemented here, which may not be ideal for such complex system with multifarious historical introgression? Rather than using these Dstat in a descriptive way, and try to interpret them, they could have been used in conjunction with hypothesis testing approach comparing formal Admixture graphs as made available in Admixture-Graph (Leppälä et al 2017) or qpGraph (Patterson et al 2012; see also ADMIXTOOL 2). Like that the authors could have formally test which graphs of species split and admixture/gene flow would provide the best fit to the D-stat or f4-stat values.

Thus, I am a bit uncomfortable with the phrasing used by the authors calling these limitations or false positive, while this look more to me like limitations on the implementation of these statistics. If the authors want to leave this like that, I would advise them to provide some perspective on alternative ways to overcome these limitations with methodological framework that make a better use of these D-stats.

This comment of providing more perspectives is actually a more global recommendation: could the authors provide more perspective on their findings, limitations, and on recent methodologies that are currently under strong development; methods capable of handling ILS and geneflow simultaneously to investigate reticulated evolution? The authors approaches here are fairly "classical" in the sense that they are using classic phylogenetic approaches (for ex. RAXML) to identify the species tree in windows, the used D stats and their derivate, which is fine. However, there is no mention in the entire document that approach such as the multi-species coalescent as implemented in BPP (Flouri et al 2020, MBE), PhyloNet (Wen et al 2018, Syst Biol), or structured coalescent (as implement in the AIM module of \*Beast (Muller et al. 2021) or even the MSMC-IM (mentioned by a previous reviewer) have been develop to tackle these limitations, and could solve in theory many of the limitations met by the authors in the present study.

This is in line with the fact that the authors are showing in this study that considering simple bifurcating tree and classic phylogenetic approaches are not appropriate to properly describe and explain the diversification processes of the Delphinoidea. Network-based approaches, MSC, or structured coalescent allowing to handle simultaneously ILS and gene flow would be a more appropriate toolbox. FYI, this is not anymore new in evolutionary biology and was already discuss at large for other biological systems like the anopheles mosquitoes, cichlids, heliconius butterflies (etc). This is why I think the authors should include a perspective

section, and discuss the case of the Delfinoidea in comparison with other well-studied system for which the same challenge has been met before. That would enrich the discussion and provide a perspective were the field should head.

Leppälä K., Nielsen S. V., Mailund T., 2017 admixturegraph: an R package for admixture graph manipulation and fitting. *Bioinformatics* 33: 1738–1740.

Tomáš Flouri, Xiyun Jiao, Bruce Rannala, Ziheng Yang, A Bayesian Implementation of the Multispecies Coalescent Model with Introgression for Phylogenomic Analysis, *Molecular Biology and Evolution*, Volume 37, Issue 4, April 2020, Pages 1211–1223, <https://doi.org/10.1093/molbev/msz296>

Tomáš Flouri, Bruce Rannala, Ziheng Yang. A Tutorial on the Use of BPP for Species Tree Estimation and Species Delimitation. Scornavacca, Celine; Delsuc, Frédéric; Galtier, Nicolas. *Phylogenetics in the Genomic Era*, No commercial publisher | Authors open access book, pp.5.6:1–5.6:16, 2020. hal- 02536475

**Answer:** The aim of our study was to investigate speciation of Delphinoidea using a suite of complementary analyses. The study is not a review of all potential analytical methods. Therefore, including a discussion of methods that are not yet widely implemented, and of results from other, unrelated taxon groups, are both outside of the scope of this study. Furthermore, because of the limitation of D-statistics when gene flow occurs between ancestral lineages, we chose to only present them in a descriptive fashion, and rather focus on the other analyses in interpretations that consider this phenomenon, namely the fbranch statistic. In light of species tree problems, we have now included a new analysis to find the species tree under the multi-species coalescent theory using Astral-III, which further supported our initial species tree based on counting the topologies of the RAxML sliding-window trees.

**209:** Same comment as above: "False positives" or just consequence of introgression in ancestors branches that lead to significant signals in the lineages deriving from these ancestors. A more nuanced phrasing would be required. See Malinski et al (2018,2021).

**Answer:** We have decided to stay with the phrasing “false positive” due to the reasoning above.

**215:** You should remain with the conditional here: I suggest "would imply"

**Answer:** Changed

**218:** Since your current phylogenetic analyses and those of McGowen et al (2020) do not really consider ILS and gene flow, and they can have a major influence on the identification of the species tree (see for ex. Muller et al. doi: <https://doi.org/10.1101/348391>), I would suggest to remain open to the idea that the species relationships for this part of the tree could be incorrect. Approach formally implementing these processes (ILS and gene flow), such as the MSC-BPP or AIM or PhyloNet are required to properly tackle these questions. The author should mention it.

**Answer:** We have now included a new analysis to find the species tree under the multi-species coalescent theory using Astral-III, which takes gene flow and ILS into account. The results are consistent with the most common tree uncovered in our sliding-window phylogenetic analysis, giving us more confidence that we used the correct species tree in these analyses.

**221:** Here would be a good place to mention some perspectives and new development capable of making a better use of these D-stats values in graphs or network. Also please mention that your current phylogenetic framework has major limitations, since RAXML is not suited to handle properly ILS and gene flow (papers such as those from Flouri et al, Thawornwattana et al) have already discussed these issues.

**Answer:** While RAXML itself does not handle ILS and gene flow, we ran this in a sliding window fashion to uncover regions showing alternative topologies to the species tree likely due to ILS and/or gene flow. We have now also included ASTRAL-III analysis to synthesise the results of the windows under the multi-species coalescent theory.

**236:** I think this is one of the most interesting analyses. However, for some reason, the authors do not provide the statistical significance of each of the comparisons. Like the D-stat and D-FOIL, the f-branch test not only provides the estimated value of excess allele sharing among the lineages and the different branches of the tree, but it also provides a formal test to assess whether the f-branch value significantly departs from the null expectation that fb value = 0.

Here, we see that all the values are incredibly small. The only exception is between the killer-whale and the ancestral branch of (Bottlenose, InFobottle), and yet here it only reaches 3.5% of excess allele sharing. All the other values are < 3%. This contrast strikingly with other systems, like the example of the cichlids where much higher values were reported (see the papers of Mallinsky et al). Therefore one could really wonder whether the excess allele sharing among the taxa along the tree are actually significantly different from 0. Could the authors provide the Z-value or the P-value, corrected for multiple comparisons as explained in the paper of Mallinsky et al 2021? Because simulations of the developers of this method show that there may be evidence of high values, just by chance, which is why a test is needed to properly assess which comparisons are worth discussing. The authors may have to revise their interpretation based on the results of the statistical test.

This would be very helpful for the readers to assess the meaning of the very small fb values.

**Answer:** We have now included the Z values for the fbranch statistic (supplementary figures S3 and S4). However, the default of the fbranch in Dsuite is only to consider fb statistics with a p-value < 0.01, which we now mention in the methods.

**239:** With respect to Fig. 2, in the panel (C), this is not clear at all to me what the authors call "hypothesized" from the f-branch results, while accounting for the inability to detect gene flow between sister lineages and a lack of positive means less gene flow ... Can the authors explain more precisely in the manuscript (possibly in the methodological section) how they come to this?

I would also suggest to add a star in the heatmap indicating which values are statistically significant, as in Mallinsky et al (2018). The authors could also put in the heatmap reporting the fb values a number on the fb-values they are actually discussing in the main text. This would help the reader quickly identify what results are being discussed.

**Answer:** We have added Z scores to supplementary figures S3 and S4. All fb shown were significant with the autosomal dataset. Most were significant with the X chromosome. The section under the new heading “Gene flow between ancestral lineages” provides a literal interpretation of the Fbranch interpretation, but also discusses some of the limitations of the analysis, and why we drew the hypothesised gene flow events from these results. We provide these hypotheses for future studies to evaluate, which may include additional species. To make it easier for the reader to follow, we have now added “(Fig. 2C)” in the text where we describe the reasoning behind the hypothesised gene flow events.

**244:** This could be consistent also with the hypothesis that Monodontidea are more closely related to the Delphinidae, than the Phocoenidae, as suggested by the DFOIL. You cannot really rule out this hypothesis based on the results of the classic phylogenetic methods used so far in this study or in McGowen et al (2020).

**Answer:** With the inclusion of the ASTRAL-III analysis we can now rule out the wrong species tree hypothesis.

**246:** As before, this could be indeed ancestral introgression, or wrong topology. Again these uncertainty in the species tree should be discussed

**Answer:** With the inclusion of the ASTRAL-III analysis we can now rule out the wrong species tree hypothesis.

**246:** Elevated may be overstated, since the fb values look more around 1 or 2% excess. Are they significantly different from 0? I suppose they are, but the authors should provide this information for all the values that seems to depart from 0.

**Answer:** We have now included the significance values.

**256:** These values are extremely small (<1%), I am really wondering whether these are actually significant, and whether they do represent signals of ancestral introgression, or indirect excess of allele sharing due to the fact that Beluga, Narwal are more related to the Delphinidae than the phocoenidae. These results, may very well represent a wrong topology of the species tree rather than actual introgression.

The authors have to discuss these alternative hypotheses: the species tree currently considered may not be accurate (due to the limitations that are inherent with the phylogenetic methods used to infer it); more advanced methods capable of handling simultaneously ILS and gene flow are required to tackle this question.

**Answer:** We have now included the significance values and an ASTRAL tree to determine the species tree.

**264:** Or wrong species tree! The authors really need to consider this possibility. They cannot trust blindly the currently accepted topology, especially in such rapid radiation where ILS

and introgression have strongly affected this system. Clearly, simple classic phylogenetic methods (like RAXML) will not be adequate to properly resolve the evolutionary history for this system.

**Answer:** With the inclusion of the ASTRAL-III analysis we can now rule out the wrong species tree hypothesis.

**274:** This is a possibility, but without having solved the uncertainty relative to the species tree topology, you cannot be conclusive. And this call for some word of caution, and the perspective on how to address the uncertain species tree topology with appropriate methods.

**Answer:** With the inclusion of the ASTRAL-III analysis we can now rule out the wrong species tree hypothesis.

**297:** You can stress here that the species tree topology needs to be further investigated with dedicated methods capable of inferring the most likely species tree while accounting for ILS and introgression. This step will be critical in future studies.

**Answer:** With the inclusion of the ASTRAL-III analysis we can now rule out the wrong species tree hypothesis.

**319:** Could you please remind here and in Fig. 3 concisely how hPSMC is estimating the divergence time and time when ILS/gene flow ends? The legend of the Figure 3, panel B, could provide a better such description to facilitate the understanding of how this is estimated from the cross-coalescent rate variation.

Please provide an explicit description of how this is estimated in hPSMC. Is it when  $N_e$  falls under the 50% or when the  $N_e$  value reaches 0, indicating no further cross-coalescent after that time point? Make sure this is well described in the methodology. For now I think this is not clearly stated how the end of ILS/gene flow is estimated from hPSMC, and how the threshold is fixed. Even if it is described in the paper describing the methods, it would be better to remind it here concisely for clarity of how hPSMC considers it as the end of ILS/gene flow.

**Answer:** The divergence times are those directly taken from McGowen et al 2020. In the methods section of this paper it is stated “We plotted the simulated and empirical hPSMC results to find the simulations with an asymptotic increase in  $N_e$  closest to, but not overlapping with, the empirical data. The predefined divergence times of the simulations showing this pattern within 1.5x and 10x of the pre-divergence  $N_e$  were taken as the time interval in which gene flow ceased.” We have also added a brief description of this in the Fig 3 legend.

**422:** I think I already mention this in a previous round of revision, but I find this entire part highly speculative and circular.

- On one hand, the authors estimate  $N_e$  variation using PSMC and try to link  $N_e$  variation with the probability to hybridize which was itself derived by methods that correlate with PSMC itself (like hPSMC). D-stat, DFOIL, D-SUITE provide a test of whether introgression may have occurred, and it does obviously, but the magnitude (as seen in DSUITE), seems

small. The global picture seems to indicate that introgression was important among ancestral branches, but biased might also be introduced due to an uncertain species tree topology.

- Furthermore, assuming variation in  $N_e$  just reflects abundance is way too simplistic. Au contraire, many evolutionary processes make it illusory to try to interpret the very gross estimation of  $N_e$  variation just as a function of abundance. For examples, gross  $N_e$  variation, as estimated by PSMC, might reflect changes in population connectivity, abundance, ancestral structure, but importantly also the various types of selective processes and linked selection, which are the neutral sites impacted by all the selective processes). For example, a recent paper by Boitard et al (2022) showed that linked selection has major impacts on pattern of  $N_e$  variation as estimated by PSMC, and thus by extrapolation by related methods like hPSMC. Thus the authors need to use these methods and interpret the results with great cautions, acknowledging all the caveats and limitations.

Thus, I strongly advise the authors to minimize this part as much as possible, keeping the speculation as minimal as possible.

Boitard S, Arredondo A, Chikhi L, Mazet O (2022) Heterogeneity in effective size across the genome: effects on the inverse instantaneous coalescence rate (IICR) and implications for demographic inference under linked selection. *Genetics* 220.

**Answer:** We agree that  $N_e$  does not just reflect population size. However, we have considered this in our discussion in several places and also discussed why or why this may not be the case for the species of interest. For example lines 461-464, 471-474, and 480-482. Although our discussion relies on speculations, which we state (line 496) we believe it is still interesting to present hypotheses about the presence of contemporary hybrids and the links with population size.

**483:** all these speculation also completely ignore the impact of the various type of selection (like background and linked selection) on the estimations made by methods such as PSMC.

The authors should really remain humble on our ability to interpret the gross global variation of  $N_e$ , and recognize that the multifarious process impacting the genetic diversity may not simply reflect change in abundance and in connectivity.

**Answer:** We considered alternative explanations than simply abundances in our results (see above response).

**524:** I do not think the problem is actually high level of divergence. Species in this super family are not that divergent. The actual problem is in fact that these species underwent a rapid diversification, thus implying short internal branch, where ILS is likely pervasive, and where introgression among ancestor was also extremely likely and multifarious. All these processes combined introduce a very high complexity making the identification of the most-likely species tree a huge challenge. Identifying such species tree topology is required to properly estimate the fraction of ILS and introgression in explaining shared polymorphism among species.

I suggest the author to consider this angle for their conclusions and also offer perspectives on how to address these challenges in future studies. More elaborated phylogenetic and population genetic approaches capable of accounting simultaneously for ILS and gene flow to estimate confidently the evolutionary history, and identify the most-likely species are required; some approach are already available such as PhyloNet, MSC of BPP, or even the newly developed Approximate Isolation with migration model. All these methods have been developed to fill the gaps left by classic phylogenetic methods. Once we are confident about the species tree, then it becomes easier to infer the most likely event of introgression, their magnitude, their timings.

Multiple genomes per species would also be needed to improve the resolution of the study, assess the importance of introgression and their evolutionary consequences.

**Answer:** We have added some text regarding challenges with short branches, ILS, and gene flow, as suggested. The new ASTRAL tree has filled in the species tree question. We prefer to discuss the conclusions in regard to the current study, and put forward hypotheses that can be tested with more genomes and analyses in future studies.

**527:** I find this point that studying sex chromosomes would be enlightening in terms of the evolutionary history of species, and the evolution of reproductive isolation, and yet the authors decided to leave these out of the present study without really justifying why. Could the authors justify the reason why leaving the sex chromosome out of this study at the end of the intro or beginning of the results, because it seems counterintuitive to me.

**Answer:** We included the X chromosome in the fbranch statistics test, as this provides the clearest result out of the tests we implemented. Further analyses on sex chromosomes are outside the scope of this study, but could be an avenue for further study, as more population-level and high-quality genomes become available.

**545:** Could the authors provide some more details of what this means in practice, and justify?

**Answer:** Liu et al showed that when the outgroup is mapped to an ingroup reference genome, it artificially looks more like the reference genome due to more “reference genome like” reads mapping successfully. This in turn biases gene flow estimates to seem like gene flow between the reference genome species and the outgroup species. We have added some more explanation about the reference biases at lines 553-554.

**553:** When we study speciation, we expect that sex chromosomes may play an important role in the evolution of reproductive isolation. Sex chromosomes may thus reveal a peculiar evolution compared to the autosomes. Can the authors justify why they have excluded them in the present study, and thus choosing to ignore that part of the genome that could be paramount to describe the speciation process?

**Answer:** We included the X chromosome in the fbranch statistics test, as this provides the clearest result out of the tests we implemented. Further analyses on sex chromosomes are

outside the scope of this study, but could be an avenue for further study as more population-level and high-quality genomes become available.

**639:** The authors can be more precise in stating which version they have used, was it 0.42 or 0.43?

**Answer:** Done

**641:** Please add the statistical test that D-suite provide (Z-score or P-values derived from the Z-score, corrected for multiple comparison). Since the values are extremely small, reporting this information is paramount to assess whether there are sufficient evidence to interpret the f-branch values. Some of these values may not be different from 0 at all.

**Answer:** Done

**642:** I have to admit I find it surprising that the authors decide for one analysis to report the sex chromosome, but not for the other analyses.

I recommend the authors to justify why they are doing that and also to explain their rationale and remain consistent in their decision.

**Answer:** We included the X chromosome in the fbranch statistics test, as this provides the clearest result out of the tests we implemented. Further analyses on sex chromosomes are outside the scope of this study, but could be an avenue for further study as more population-level and high-quality genomes become available.

**656:** reference malformed

**Answer:** Corrected

**753:** I personally do not understand this explanation. Could you please rephrase it, and also provide more details in the methodological section?

**Answer:** The details for this are best explained during interpretation, which we have specified in the main text under “Gene flow between ancestral lineages”. The figure legend is to be read while reading the results/discussion section to fully understand the decisions we made behind these hypothesised gene flow events. However, we have now made it clearer that we form these hypotheses rather than the method itself doing so.

**758:** Please provide an explicit description of how this is estimated in hPSMC. Is it when  $N_e$  fall under the 50% or reach 0? Make sure this is well described in the methodology. For now I think this is not clearly stated how the end of ILS/gene flow is estimated from hPSMC, and how the threshold is fixed. Even if it is described in the paper describing the methods, it would be better to remind it here concisely for clarity of how hPSMC consider it as the end of ILS/gene flow.

**Answer:** We have now added more description about the hPSMC method to the figure legend.