Dear Dr. Westbury and colleagues,

After reading again the manuscript in light of the previous rounds of revision, I am still convinced that this work is of great interest, but there are still some issues remaining that need to be addressed.

I have decided to ask for a fresh look from an expert in the field of reticulated evolution, Dr. Simon Martin and I have provided also detailed feedbacks incorporated in the attached version of the manuscript. Dr. Martin's review and feedbacks encapsulate many of the remarks and issues I have on this manuscript, but I do have also some specifics of my own (in green in the pdf). Please check the reviewer's feedbacks/comments and mine in the annotated pdf version.

I would advise the authors to follow these suggestions. I think they would benefit greatly to the manuscript and make it clearer and more accessible.

I am looking forward to receiving the revised version in order to complete the recommendation in the PCI Evol Biol.

Best regards

Michael C. FONTAINE

Download recommender's annotations (PDF)

Answer: We would like to thank Michael for taking the time to comprehensively assess our manuscript as well as include a new reviewer. We believe that the suggestions made by both, especially the inclusion of the Fbranch tests, have greatly improved our interpretations and strengthened our hypotheses. We have answered all comments, which we include in this document, to the best of our ability and hope our manuscript is now more suitable for a recommendation.

Kind regards (on behalf of all coauthors),

Michael Westbury

Editor comments

Abstract Half of the abstract provide general context, while the remaining is rather superficial and a bit unspecific, there is no specific report about which hypotheses / questions being addressed, and the main findings of this study, and the main conclusions.

You may want to revise the abstract accordingly.

Answer: We have revised our abstract, with less emphasis on context and more on what this study is specifically about. The abstract now also includes a brief sentence on the results from each of the main method classes: phylogenomics, shared derived alleles, and demographic.

40 which barrier are you reffering to? The "these" imply that you have presented the barriers in the previous sentence, but this is not very clear. You may want to rephrase the previous sentence with replacing "isolation" by "barriers".

Answer: Changed barrier to isolation

49 these are cetacean specific references, if I am not mistaken. you could precise this in the sentence, or include a more general review about this, may be?
Answer: We have now added that gene flow has been uncovered using genomics in distinct taxonomic groups to make it clearer it is not only about cetaceans. The references we cite include baleen whales, bears, and hyenas, which we believe reinforces our comment about gene flow being found more widely with new genomic data availabilities.

52 I would put "but usually requires". because there may be many exceptions to this general statement.
Answer: Done

67 The reference of Andy Foote may be suitable here.
Foote A.D. (2017) Sympatric Speciation in the Genomic Era. Trends in Ecology & Evolution.
Answer: Added

120 you may want to explain why this could be the case. If the topology discordance is caused by interspecific geneflow, then I agree that the more ancient the event is, the smaller the block with a discordant topology would be. However, under ILS, this is not necessarily true. So you may want to be more precise in your statement.

Note also that the new reviewer has made a usefull suggestion in this respect, to assess wether introgression is recent or ancient. This may be a very usefull addition to the present paper.
Answer: We have now specified we are talking about gene flow events. We have also run the suggested analysis, which confirmed our hypothesis of ancient events.

134 Note that there is not much explanation on how to read the QuIBL results. You need to provide more explanations, and better describe the results and their meaning. This is valid for Table S3 and Table S4.
Answer: We have added much more interpretation to the main text about the QuIBL results as well as more information to the supplementary table legends, and believe the results are easier to interpret now.

138 What results are actually suggesting that? Not that they are not, but there is just no description of the results. Readers may not be used to QuIBL and how to read these results. The authors must make a better job a describing them, for a non-specialist audience. This comment is valid in general for the this analysis and the others as well.

Among which species is this the case that a model of ILS+gene flow provide a better fit? Make sure also to cross these results with the other tests as well. Are all the analyses you have run (QuIBL, D-stat, etc) concordant or not?

Answer: We have added more detail to this section as well as to the table legends, which clarify the QuIBL results. We have now added a statement at the end of the 'detecting gene flow' section about the concordance of the different tests for gene flow.

146 Like the reviewer, I was surprised that you have found that all tests suggested evidence of gene flow. All the tests but one are significant.

Could it be due to a wrong species tree used, or violation of some underlying hypotheses in this test, other issues, or biases? Durand et al 2011 (MBE) and Green et al (2010), the original author of the ABBA-BABA test states that, to tease apart gene flow from ILS, the pair H1 and H2 should be more closely related in general than the test species H3. In your case, here you test all the permutations.

Also here, introgression signal may be on the ancestral branch rather than on the species under focus. For example, you do not suggest there was actual gene flow between the narval and the finless or the harbour, but with their ancestors, is that correct?

You should thus provide a better explanation why all the D-tests but one are actually significant, and the meaning of this, plus potential bias, limitations and issues.

See also the comments of the reviewer and other comments later in the MS about this issue.
Answer: As we recalculated the species tree through a sliding window + consensus approach in this manuscript itself, it is highly unlikely that the species tree input is incorrect. While we test all permutations, we also discuss later on how this could bias results including ancestral gene flow events so chose not to discuss the comparisons in depth too much due to potential overinterpretations. The potential for D-statistics overinterpretations is why we chose to also include the Dfoil and hPSMC test. However, now with the addition of the Fbranch results, it seems clearer that our high signals of gene flow between species tested here are more likely to be the result of more ancient ancestral gene flow events which we hypothesised when discussing all results.

152 reference paper describing this method?
Answer: Done

158 can you be more specific?
Answer: Done

165 Please provide a better explanation on how to read these tables in the caption.
Answer: We have now added more detail into what each part of the table means to aid in interpretation.

175 here is the possible explanation. May be add a note on line 148 to explain how and why could the D-stat misleading.
Answer: We have kept wording as is as this potential problem is only visible in the Dfoil results, not in the D-statistics results alone. We have added that if this was the case the significant results would be due to more recently shared ancestry than gene flow.

176 note that if there has been so much ILS and geneflow, as we expect in Delphinoids, it may be likely that the species tree as retrieved in these previous study may be wrong. A classic concatenation phylogenetic approach may be poorly able to retrieve the correct species tree if ILS and gene flow is important as shown by Thawoenwattana et al (2018, MBE). And multi-species coalescent phylogenetic methods that ignore gene flow can also lead to the wrong species tree as shown by Müller et al (in press, Syst. Biology).

Thus you should not want to rule out too quickly that the species tree is not incorrect, especially in your study where you show that reticulation is common among cetacean species.

Thawornwattana Y., Dalquen D., & Yang Z. (2018) Coalescent Analysis of Phylogenomic Data Confidently Resolves the Species Relationships in the Anopheles gambiae Species Complex. Molecular Biology and Evolution, 35, 2512–2527.
Mueller N.F., et al (2021) Joint inference of species histories and gene flow. Syst Biol, in press and bioRxiv preprint doi: https://doi.org/10.1101/348391

Answer: Yes we agree that using phylogenetic methods that ignore discorandances can be problematic. However, as we needed a fixed topology as input for our analyses, we decided on the most common tree from our sliding window phylogenetic analyses, assuming that is the species tree and all other topologies arose due to ILS/gene flow. It is important to note that we not only used the topology from previous studies using smaller datasets, but also confirmed that this was indeed the most frequent topology on a genome-wide level.

186 Did these studies actually show that D-stat and D-Foil could be biased by these factors or is this your opinion that these statistics are important and may biais the results? Please verify that the reference cited actually showed that these effect could bias the D and DFOIL test, and adjust your statement.

I would still let open the door that the species tree as infered here and previous study may require revision or re-investigation, given the fact that they ignore the potential importance of gene flow in the resulting species tree.

See Mueller N.F., et al (2021) Joint inference of species histories and gene flow. Syst Biol, in press and bioRxiv preprint doi: https://doi.org/10.1101/348391

Answer: Yes, these biases were uncovered through simulations in the cited papers.

207 which supplementary results are you reffering to?
Answer: Here we are referring to the link in the supplementary information under the supplementary results section. We have now changed this to supplementary data - hPSMC to improve clarity.

208 I may be mistaken, but did not McGowen and coll used also multi-species coalescent approach to estimate their trees, thus taking into consideration ILS, but still not gene flow?

If I am mistaken, and indeed McGowen did not used MSC approach, then my comment above about ruling out too quickly the wrong species tree to explain the issues with D and DFOIL is thus even more likely.

Answer: McGowen et al used a consensus tree to estimate their times, which did not take discorances into account. In regards to the wrong species tree being used by McGowen et al, as mentioned above, we recalculated the most frequent topology using various sliding windows and these were congruent with McGowen et al.

221 For sake of clarity, can you remind that this is the hPSMC results here that suggest that.
Answer: Done

241 did you combined autosomes and X-chromosome in this analysis, I think this is juse autosome, correct? Of course comboining the two would be a poor choice, given the difference in inheritence.
Answer: Yes this analysis was only using the autosomes.

Note that the X-chromosome usually evolve faster and accumulate more rapidly reproductive isolation factor involved in the speciation. Did you try to run hPSMC only the X-chromosome to see if you get a different picture, faster sorting of the lineage and also lower gene flow?
Answer: We only used the autosomes and have clarified that at the top of the segment. As we were interested in the overall speciation process in Delphinoidea, we focussed on the autosomes here. However we have now added the Fbranch test also with X chromosome scaffolds only, which show the same overall pattern as the genome-wide dataset.

256 make sure the caption are complete and accessible. Currently S3 and S4 are not.
Answer: Corrected

277 I have the impression that ancestral population structure could also create this? Indeed fragmented populations could fix randomly distinct haplotypes in the distinct populations, and be maintained. Please check alternative to just the very large Ne.
Answer: We were unable to find any specific literature on this. However, one could argue that ILS cannot occur without population structure. Before a speciation event, populations would need to become structured/isolated from one another and it is only after this isolation from one another that ILS can occur. In this sentence, we are specifically referring to the long duration of ILS/gene flow that we observe after initial divergence, which could be explained by large Ne and therefore a lesser impact of drift on the loss of alleles inherited through ILS.

291 I may want to insist here on close relatives. Your analyses suggest that hybridization may take place also between non-closely related species.

Answer: Changed to just relatives as hybridization can occur between more distantly related species too

306 PSMC estimates variation in effective populations size. I insist on the effective, here. The link between effective and census size is extremely difficult to make, and most if not all studies that tried to link the two usually failed or came with very strong assumptions. There are so many determinants impacting Ne variation, such as ancestral structure / population connectivity, introgression / geneflow, selection, and so on. Thus, since introgression can have a strong impact historical Ne, using Ne to infer whether a species is subject to introgression looks completely circular to me. There is a fundamental problem of circularity here, because a species receiving gene from closely related species would see its genetic diversity increasing, and thus its Ne also increasing. This is thus a fundamental issue in the rational here, effective (Ne) is not census (Nc). I would personally suggest to drop this part or to reduce it to hypotheses.

The reviewer suggested a method to assess whether introgression was recent or ancient. I suggest you dig into this first. Then there is also way to date when introgression occured. **Answer:** We have significantly reduced this section. However, the PSMC results we discuss here are between sister species, which the other analyses are unable to infer gene flow between, meaning that we can only discuss the hPSMC in this context.

325 but, it is very difficult to use here Ne as a proxy of census abundance here, since Ne is an abstract measure of genetic diversity, and the evolutionary forces that act on it. Ne will thus reflect historical variation in population size, mating system, population structure, gene flow/introgression, and selection.

The interpretation made here, are too simplistic, circular, and speculative. **Answer:** We have added an additional explanation as to why our Ne results may truly reflect abundance rather than be the product of gene flow.

359 All in all, I find this entire section speculative. Indeed the circularity of using Ne is problematic to me. This is a measure that is itself impacted by all the evolutionary forces (drift, geneflow, and selection). Thus using it as a measure of abundance seems inappropriate to my point of view. I would suggest be reduced this entire section to its minimum and just keep this as an hypothesis. **Answer:** We have now shrunk this section and hopefully explained our hypotheses better.

380 this is especially true in cetaceans, since their genome is known to be extremely conserve, stable and evolve slowly. Thus incompatibility may thus arise very slowly too, whcih means hybrids would remain viable, long after the species diverged. **Answer:** Yes, this is why we present parapatric speciation as a major method of speciation in our study system.

410 such as? you should some words of caution about potential biases that can arise, due ton the fact of using a very divergent reference genome. I think a previous reviewer (Andy Foote) made comments about this.

Answer: We have now added more explanation and changed the citation

416 how? please be more precise.

Answer: Using a synteny based approach with the software Satsuma as clarified in the text now.

417 this is actually a pity that you did not analyze the X chromosome alongside (but separately) to the autosomes. It is well known that sex chromosome evolve faster, and incompatibilities creating reproductive isolation accumulate primarily on the sex chromosome. Thus in many taxa, sex chromosome contribute disproportionaly to the speciation process. You should discuss this limitation in the main text.

Answer: As the Fbranch results gave us the clearest and easiest to interpret results as to gene flow events, we decided to also run this analysis with the X chromosome only. The results from this are very similar to the genome-wide dataset and we have discussed the potential causes of the differences in the main text. Moreover, we have added some information about the possibility that different regions of the genome may show different contributions to reproductive isolation in our conclusions, and that our genome-wide picture is the first step to understanding the process in greater detail.

421 please explain why

Answer: Done

450 which results? Be more explicit, please.

Answer: Added "discordant phylogenetic topologies"

473 please add the reference of the test. also at the population level, it make sense to test all the combination. On the other hand, testing all possible triplet of species may not make sense at the macro-evolution level. Should not you restrict your test to consider H1 and H2 to be closely related? Otherwise, it might violate the assumptions of this test.

485 same comment as for the D-stat: all possible combinations may not make sense, because they would violate the hypothesis of this test.

Answer: We have now added a reference. Yes, this may violate the assumptions which is why we present the assumptions in the main text alongside the results from all combinations. This is why we chose to only touch upon the D-statistics results as complementary results that could support or refute hypotheses from the other analyses. We also include the Fbranch now, which was very helpful in interpreting the potentially biased results.

487 how did you asses the significance of the DFOIL test? Please specify.

Answer: The significance of each window is separately assessed by a chi-squared goodness-of-fit test within the software and we specify this in the methods now.

Figure 2: define what is AR.
Answer: Auto-correlated rate - added

Supplementary tables S3 and S4 provide more explanations and details for the reader. Also explain the title of each column. Readers may not be expert in reading these results. You need to make them accessible and understandable.
Answer: Added

S6 Add a note here where to find this table. Also please provide a much better explanation on how to read these tables.
Answer: We have added a note that it is attached as a spreadsheet and more information as to what each of the table descriptors mean..

Reviews

Reviewed by Simon Henry Martin, 31 Aug 2021 22:00
Westbury et al. present a genome-scale assessment of gene flow in the Delphinoidea. Using a combination of approaches that consider genealogical discordance and branch length, they find evidence for a strikingly abundant "post divergence" gene flow in this group, including between families. While I initially found the results hard to believe, I do think that the most parsimonious interpretation of these findings is that the history of this group is far from a simple bifurcating process, due (at least in part) to widespread gene flow causing systematic discordance in genealogies across the genome.

Overall, I think this study is a good first step towards understanding speciation and the role of hybridisation in this lineage. Nevertheless, I still have multiple suggestions for improvements to the Results and Discussion section, as I felt that it was too brief and did not adequately describe the findings and their possible interpretations. I also have one suggestion for an additional analysis that I think will help with interpretation.

1. QuIBL
The results from the QuIBL analyses are described with a single vague sentence "Our QuIBL analyses suggest that the different retrieved topologies cannot be explained by ILS alone, but a combination of both ILS and gene flow."
This is quite unsatisfactory as a description for such a broad set of tests. I would recommend adding a more detailed description of which pairs of non-sister taxa showed evidence for gene flow. Looking at tables S3 and S4, the results are striking and not immediately easy to interpret. There are so many significant signals, including support for gene flow between killer whales and every other member of the Delphinidae! I think the wording in the main text needs to capture at least the scale of significant signals, and ideally some discussion of what this would imply if the results are accurate (i.e. that gene flow might be or have been rampant between these species).

I also think the description of supplementary tables S3 and S4 requires more detail, as some of the column headings are not clear to the reader who does not know the QuIBL program.

Answer: We have now added a much more detailed description of the QuIBL results and discrepancies that arose due to window size and triplet outgroup. We have also added a greater description of how to interpret the supplementary tables to the table legends.

2. D statistics

Like the QuIBL results, the D statistic results are not actually described in the manuscript, except a general statement that "85 out of 86 tests show signs of gene flow" (line 146). Looking at Table S5, we see that the results are even more astounding than those of QuIBL. Again, I recommend adding some description and interpretation in the text. Which pairs are consistently showing evidence for gene flow? Are the results consistent with QuIBL? Could any of the results indicate some systematic problem like the tree being incorrect?

It is good that the authors acknowledge that ancient gene flow events can be difficult to distinguish from multiple more recent events. However, they do not offer any detail as to how and where ancient gene flow might explain their specific findings in this study. For example, could ancient gene flow between killer whale and the ancestor of some dolphin species explain some of the apparent signals better than multiple more recent events? I know the authors have already added several additional analyses to appease reviewers, but one more I would strongly recommend is the f-branch statistic (Malinsky et al. 2018 Nature Ecology and Evolution, https://www.nature.com/articles/s41559-018-0717-x#Sec13, https://github.com/millanek/Dsuite) which provides an explicit approach to identify ancient events that affect the descendent lineages.

Answer: Generally speaking, the results are consistent with QuIBL but with some discrepancies, which we now mention. We initially chose not to discuss D-statistics in depth because of some of the biases that may be arising due to the limitations of the method, especially when there were likely ancestral gene flow events. However, it is unlikely that an incorrect topology is the driving force behind the significant results, as we only present topologies consistent with the species tree - determined as the most common topology in our sliding window analyses.

We have now added some more information to the results, but still limit the interpretations because of the potential for overinterpretation due to unforeseen biases. Instead, as we have run the suggested Fbranch test and it has greatly helped in the interpretation of such prevalent positive D-statistics, we chose to add much greater description into the results in another section focused on the Fbranch results.

3. D-foil

The description of these results is better, as it provides some interpretation and possible caveats. However, the final statement "we suggest our result reflects the limited ability of D-foil to infer gene flow between these highly divergent lineages." needs more explanation.

Answer: This final sentence was meant to link into the next paragraph, but we see now that that was not clear. We have reworded this to tie these together better.

4. hPSMC

The hPSMC results suggest that gene flow continued until within the past 5 million years between all species considered. These times are difficult to reconcile with the divergence times estimated by McGowan et al. 2020, which are between 10 and 20 million years for most species. The authors appear to interpret the McGowan estimate as the "divergence time" and the hPSMC estimate as the end of "post-divergence gene flow", but I think this reasoning needs to be justified by explaining why the McGowan estimate might be less impacted by post-divergence gene flow.

Answer: McGowen et al used a predefined tree as input for their calculations, which we now mention in the paper. This means that discordances (ILS or gene flow) are not unlikely to be captured in their estimates. Furthermore, fixing a topology despite the presence of gene tree discordances is known to extend terminal branches, because molecular substitutions of discordant loci need to go somewhere (Mendes, F. K., & Hahn, M. W. 2016, Systematic Biology, 65(4), 711-721.) which will systematically lead to older node ages, and may be the case with the estimates by McGowan et al. We have also added a new paragraph about why we may be getting some of these longer post-divergence gene flow periods, by mentioning how the genomic remnants of ancestral gene flow events may be differentially segregating due to ILS; which may explain some of the seemingly unrealistic hPSMC results.

5. Tree node age analysis

The node age analysis is a nice addition to the paper, but the authors should acknowledge that it is not really independent from QuIBL. Arguably, QuIBL does a better job because it models the expected mixture of two different node age distributions resulting from ILS and gene flow. Nonetheless, the tree-based approach has appeal due to its simplicity, and the fact that we can consider multiple nodes at once.

I would also like to see some explanation for the choice of the three focal species used. Specifically, why were the two bottlenose species not included despite also showing ample evidence for gene flow in the other analyses?

Answer: We selected these comparisons as they had the most loci showing discordances and as the more loci we could include in the analysis, the more accurate and balanced our comparisons of relative dating results would be. We have included this reasoning in the methods section. Although there are some parallels with the QuIBL method, we believe we still get some extra information compared to the QuIBL results. QuIBL simply uses the different branch lengths to determine gene flow/ILS but does not give insights into when the gene flow may have occurred.

The reviewer is correct that our inferences of relative divergence times may be pulled to older dates due to the inclusion of loci with discordant topologies that arose due to ILS instead of gene flow. As it is very difficult to differentiate signals of ILS and gene flow, we chose more conservative wording when presenting the results by stating the results reflect both ILS and gene flow. We have added a summary at the end of the section about how discordances arising due to ILS and gene flow would present different relative timings.

6. Geography

There is quite a bit of text about allopatric vs parapatric speciation, but there is nothing about the distributions of these species, how much they overlap, and whether parapatry/allopatry is/was likely. Some information about their distributions will aid interpretation and could also help set up some predictions at the start of the paper, which would make it less descriptive.

Answer: As we are talking about ancient gene flow events (now further confirmed by the Fbranch results), we thought that adding descriptions of modern distributions may lead to false conclusions being drawn about where gene flow events may have occurred. Due to their aquatic lifestyles, the fossil record for cetaceans is relatively scarce and therefore, we have very limited information about distributions millions of years ago when we hypothesised gene flow events to have occured.