

Dear authors,

The manuscript entitled “Speciation in the face of gene flow within the toothed whale superfamily Delphinoidea” by Westbury and colleagues has been now evaluated by three referees. They all agree that this study has great merit and should be of interest to the community. However, the three reviewers also raised major issues and concerns that I am also sharing.

In particular, the three reviewers acknowledged that the phylogenetic analyses were properly conducted, the interpretation from the results were sounds and the limitations properly acknowledged. See however the comments made by the three reviewers, especially reviewer 1 about incongruences in tree topologies compared to the consensus tree, which I am also sharing. The authors should also check the recently develop approach "QUIBL" developed by Edelman and colleagues (2019; Science: 10.1126/science.aaw2090) to test for introgression without the limitation of symmetric tree topology imposed to the DFOIL test. This would be complementary to the test currently presented in the manuscript.

Regarding the second part of the analyses which aims at assessing and modeling post-divergence gene flow based on the use of TMRCA variation along the genome inferred from hPSMC, the three reviewers raised serious concerns and all agreed this part need improvement and I agree with them. This is especially important, since the title relies strongly on this part, and yet the results supporting this "speciation with gene flow" deserve further attentions and validations. The three reviewers made very good suggestions to explore the data further, interpret the results and discuss better their limitations. These would improve the study greatly.

I strongly suggest the authors to consider all the reviewers' recommendations.

I would be happy to reconsider this manuscript for a recommendation in PCI Evol Biol if the authors can continue to develop this study further and address point by point the three reviewer's comments.

Best wishes and already Happy New Year.

Michael C FONTAINE

Ref. Edelman N. B., et al. 2019 Genomic architecture and introgression shape a butterfly radiation. Science 366: 594–599.

**Answer:** We would like to thank Michael for taking the time to evaluate our manuscript and for bringing QuIBL to our attention. It was a useful tool. We have now additionally run this analysis on two different datasets, the 50kb windows from our original phylogenetic analyses, as well as a new dataset using 20kb windows in an attempt to assess as much ancient discordance as possible. In both of the datasets, this analysis has confirmed that ILS alone is not enough to explain the observed patterns. However, since we only got alternative topologies within Delphinidae, we were only able to test within this family. Despite this, we believe that our results are greatly strengthened with this new addition.

## Reviews

### *Reviewed by Andy Foote, 2020-11-26 11:12*

Westbury et al. investigate the evidence for divergence with gene flow among toothed whale species making use of the increasing number of reference genomes. This is a worthwhile venture and given how prevalent introgression and hybridisation are becoming in genomic studies, and the cases of known hybridisation in this taxonomic group, it seems a tractable question to pursue.

I should declare at this point that I have known and worked with Eline for many years and so there is a potential conflict of interest. However, I discussed this with the editor and we agreed I could give an impartial review.

**Answer:** We would like to thank Andy for taking the time to look through the manuscript so thoroughly and provide us with a lot of thoughtful comments. We have addressed all comments below and believe that our analyses and therefore conclusions are more robust because of them.

To identify potential signatures of post-divergence gene flow, the authors first build a phylogenetic tree and look for topological incongruence among windows of 50 and 100 kb. They then perform statistics (D-statistics & D-foil) that infer gene flow from excessive derived allele sharing compared to that expected from the hypothesised tree topology. Lastly, they look at variation in TMRCA across the genome between a randomly sampled haploid genome from each of two species using PSMC. They then compare when they detect a cessation of coalescence with when a recent published study estimated divergence times. The conclusion of speciation with gene flow mainly stems from this last set of analyses.

The difficulty with this type of comparison is that the study species mostly diverged millions of years ago and many of the inferred introgression events also occurred in this time frame. Therefore the length of any introgressed tract will have been broken up by the action of recombination and the task of teasing apart introgression, ancestral population structure and incomplete lineage sorting becomes challenging. And secondly, it makes comparing across such divergent genomes essentially a comparative genomics study rather than a population genomics study.

Regarding the first difficulty, I feel the authors do a good job of treading conservatively when interpreting the results of their phylogeny, D-statistics and D-foil results. They do not strongly infer one process over another, which I think is a good call. They do however come down in favour of an interpretation of divergence with gene flow from their pseudo-diploid PSMC plots, which I feel needs more scrutiny by the authors (see comments below).

Regarding the second difficulty, the authors adopt a more population- than comparative-genomics approach: mapping short read data from each study species onto an outgroup reference genome (the baiji). Whilst this removes reference bias, mapping data to such a diverged genome brings up potential issues of chromosomal rearrangements, which are especially problematic when contiguity is assumed in the analyses, for example, the hPSMC analyses. At the minimum I would have liked to have seen syntenic maps comparing among

the reference genomes to understand if there are some problematic regions which should be excluded.

**Answer:** We understand the reviewers concerns, and have addressed this point in a previous publication (<https://doi.org/10.1371/journal.pone.0222004>), where we mapped orca whale genomes to either an ingroup reference genome or a highly divergent outgroup species (bowhead whale). While there was some potential reference bias for the D-statistics analyses when mapping to the ingroup reference genome, we showed that hPSMC was unaffected and produced the same result regardless of reference genome. We have now added a sentence about this to the hPSMC methods section.

Further, although the authors do apply stringent mapping QC measures, they don't appear to explicitly mask repeats or other regions of potential low mapability. These make up around 45% of the killer whale genome for example. These regions of the genome could generate heterozygous calls that could influence all the analyses conducted in this study. I would recommend the authors curate a set of homologous regions of sufficient length to be included in the analyses that assume contiguity.

**Answer:** We have now rerun three hPSMC using species pairs of various divergences (shallow - bottlenose and indopacific bottlenose dolphins, medium - narwhal and beluga, and deep - beluga and the bottlenose dolphin) and removed the repeat regions presented in the genbank river dolphin annotation. The results of this are now found in supplementary figures S2-S4 and are mentioned in the main text. While there was a slight difference between the original and these results, the pre-divergence  $N_e$  looks almost identical, and the exponential increase in  $N_e$  is not dramatically different from one another, meaning that our results should not be greatly influenced (if at all) by the inclusion of repeats. Rather, this may reflect our strict filtering, as well as the method we implemented randomly selecting a consensus haploid base here, not taking heterozygosity into account meaning that if there were erroneously called het sites due to repeats, this is unlikely to greatly influence the final hPSMC result.

The authors state that the RAxML phylogenetic tree showed consensus support for topology in the McGowan et al. 2020 paper for all window sizes. But they state when they use a window-size of 50 kb, up to 76% of windows are incongruent to this consensus. I would like more details of this incongruence. Does this result mean that the congruence is driven by just 24% of the windows across the genome? Presumably, the 76% showing alternative topologies showed an array of different topologies? This seems to me like an expectation from using such a small window size (how many SNPs were there per window at 50 kb?) and potentially from including regions of the genome that should be masked thereby potentially comparing among prologues (although the authors do stipulate they only use uniquely mapped reads).

**Answer:** Yes, we observed an array of different topologies, which we hoped could be retrieved with our Densitree plots. However, now we see that maybe this was not stated in a clear enough manner, and have included a supplementary table with (i) the 5 top topologies per window size,

and (ii) the proportion of windows for which each topology occurs (Supplementary table S1). Also, we have included the number of unique topologies per window size in the main text. This variability may arise due to differences in the number of SNPs per window. However, upon further investigation, we saw that the minimum number of variable sites for a 50kb window was 899, which should still harbour a lot of information. Additional values for variable sites are maximum - 4,588, median - 2,248 and mean - 2,301. As for the potential problem with comparing paralogs, we mapped the raw reads from all species to the baiji and built consensus sequences from that, so they would all be a single homologous alignment. Furthermore, just in case there were mapping problems caused by repetitive elements our strict filtering options should have reduced the majority of biases.

Regarding the hPSMC method. I feel the pseudo-diploid approach to estimate divergence times using phased haploid markers such as male X-chromosomes (see Li & Durbin 2011) is helpful in population genetics studies. I am more sceptical of randomly sampling two haploid genomes from diploids to create a pseudo-diploid. This will underestimate heterozygous genotypes, for example at sites where both individuals have heterozygous genotypes there will be a 50% chance of sampling a homozygous genotype. So that for example argues against the statement "The variation in the interspecific pseudo-F1 hybrid genome cannot coalesce more recently than the emergence of reproductive isolation between the two parental species, and the method can therefore be used to infer when gene flow between species ceased."

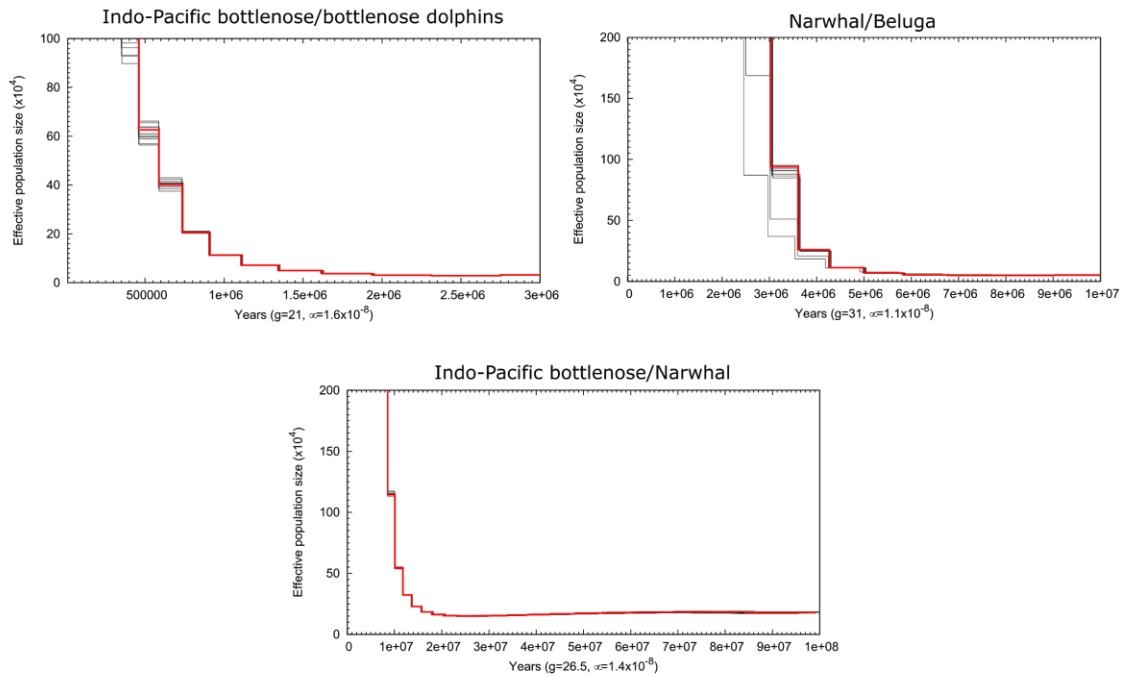
**Answer:** The hPSMC analysis is concerned with fixed differences between species, as opposed to between populations. As the reviewer mentioned, in theory if the site was heterozygous then there is 50% likelihood to select each allele across the entire genome. This may cause issues if the number of heterozygous sites was high. However, the fixed species differences should completely swamp any additional differences that may be added (or removed) due to heterozygosity in the species. We compared our results to see if this was the case. However, even when considering the species with the smallest divergence (Indo-Pacific bottlenose and bottlenose dolphins 0.004) and the highest heterozygosity of the pair (Bottlenose dolphin ~0.0003), there is still an order of magnitude between them, which is likely to drive the result based on the fixed differences, as opposed to the occasional incorrectly called heterozygous position.

Although the Cahill et al. paper estimates from ms simulations that hPSMC can approximate divergence times up to 5 MYa from unphased pseudo-diploids, these simulations did not incorporate the true complexity of actual genomes in which selection, mutation rate and recombination rate vary across the genome. The inference of divergence with gene flow in this preprint is based on comparing the divergence time estimates from McGowan et al. with those inferred from their hPSMC analyses. But McGowan et al. used homologous exons, markers likely to be under selective constraints, and so perhaps it is not surprising to see that some differences in coalescent times between the two inferences? Could these differences in TMRCA be explained by variation in these key parameters across the genome?

**Answer** - Although there could be some slight differences if a different divergence date data set was used (when we plotted the hPSMC we used a mutation rate that was calculated based on the estimates from McGowen et al), our resulting conclusion will not change, as the exact dates are somewhat arbitrary. This is the main reason why we simply state that gene flow continued for a long time after initial divergence, and report relative lengths of the branch, as opposed to absolute lengths of time. If we had used an older or younger divergence date to compare against our hPSMC results (accordingly, the mutation rate would then have been either faster or slower), the relative difference would not have changed.

Lastly, the pseudo-diploid PSMC plots show changes due to the rate of coalescence at different time points, inferred from tract lengths of homozygous and heterozygous 100bp bins. Incomplete lineage sorting can be the cause of shared genetic diversity a long time after species divergence, so that for example, under a simple allopatric speciation model, drift alone requires  $\sim 10 N_e$  generations to make incipient species reciprocally monophyletic at more than 95% of loci (Hudson and Coyne, 2002). So that species with long generation times as is the case for the whales and dolphins in this study, often share genetic variation for many thousands of years after divergence. Assuming  $N_e$  of 10,000, based on killer whale data, that would equate to 100,000 generations, or 1.5 million years. Note that the length of such shared tracts either due to ILS or gene flow dating back to  $\geq 1$  million years ago would be small due to the action of recombination, that artefacts creating wrongly diagnosed heterozygous or homozygous 100 bp bins could impact the analysis. It would be useful to see bootstraps of the hPSMC plots just to see how noisy these plots are.

**Answer** - The hPSMC analysis uses 10bp bins, as opposed to 100bp, to better capture the variation found between species as opposed to within species (heterozygosity). If 100bp bins were used, then simply due to the larger differences between species (relative to what is found when considering two chromosomes from a single species), a very large proportion of bins would have had at least 1 heterozygous site, which would heavily influence results. However, to test this we ran an additional 10 bootstraps for three different species pairs representing deep divergence (Indo-Pacific bottlenose dolphin/narwhal), medium divergence (beluga/narwhal), and more recent divergence (Indo-Pacific bottlenose/bottlenose dolphins) to assess the variance of our results. We find that the bootstrap values are very similar to those produced with the whole dataset, suggesting incorrectly diagnosed bins is likely not a large issue with this analysis. We have included the figure below in the reviewer replies showing the results of the bootstrap test. The red line shows the whole genome hPSMC result and the black lines are the bootstraps.



In summary, I feel while the authors did a judicious job of highlighting that either incomplete lineage sorting or gene flow could explain many of their results, I suggest the same could be said for the hPSMC analyses and the same caution should be expressed for these results. Further, I think methodological artefacts may also play an under appreciated role and requires some further data exploration. Taken altogether, my comments and the authors' own reflect the difficulty in inferring the underlying processes in such ancient shared variation. Perhaps that might be a better and tractable central message for the authors to focus on? I feel this would be a very worthwhile outcome, given that many such studies are now finding what they infer as signatures of introgression and hybridisation. I hope these comments are helpful and that the authors are able to continue to develop this study, and I wish them all the best with publication of a suitably revised version.

**Answer:** Based on the reviewers suggestions, we have now added a closing remark about the difficulties of resolving such complicated questions with limited data sets but that general patterns can be assessed when including a multifaceted approach.

*Reviewed by Christelle Fraïsse, 2020-11-26 16:29*

Westbury and collaborators present an interesting study investigating post-divergence gene-flow in toothed whales. I liked the multifaceted approach combining phylogenetics, descriptive summary statistics and demographic models to tackle this question. The text is clear and concise, interpretations of the results are generally wise, and analyses well-conducted. I am convinced that this work will be a great addition to the speciation genomics literature, in particular because it highlights the unusual features of the marine realm and their consequences to speciation. I only have one main comment, and a few suggestions to improve the paper.

**Answer:** We would like to thank Christelle for taking the time to look over our manuscript and for her suggestions. We have carefully considered each point and provided responses to each point.

#### MAIN COMMENT

My main point is that the demographic analyses of post-divergence gene flow (L168-210), and its relation to species abundances (L212-287), could be improved.

I understand that the authors “only” have in hand a single genome per species to reconstruct their demographic history, and so this limits the type of demographic analyses that can be performed. In the paper, the authors employ the hPSMC method that uses a pseudo-diploid F1 hybrid genome to deduce the time of cessation of gene flow between species. This time is indirectly inferred based on its effect in reducing the estimate of divergence times. Then, to test for the influence of species abundance on hybridization, the authors estimate population sizes through time using PSMC, and discuss the two features (post-divergence gene-flow and species abundance).

- First, I think that the connection between the two features could be made more readable. For example, you could add on Figure 3B,C,D the corresponding intervals for the cessation of gene-flow.
  - **Answer:** We have considered adding the end of gene flow results to the PSMC plots, however we have decided against it as we would need to incorporate information from 10 pairwise comparisons in the Delphinidae plot (Figure 3B), which would completely swamp the figure and decrease its overall readability.
- Second, as acknowledged by the authors their approach has some limitations, including that i) it does not provide the direction and rate of post-divergence gene-flow and, ii) it cannot disentangle a decline in species abundance vs. a reduction in interspecies gene flow. That is why I think it would be worth trying the Sequential Markovian Coalescent method implemented in MSMC-IM (Wang et al. 2020: doi:10.1371/journal.pgen.1008552). This will help to better evaluate the recent demography of each pair of species, including getting an estimate of both population sizes and migration rates over time. One difficulty is that this tool requires phased data, and so I am not sure whether the authors will have the relevant data for all species. Anyway, applying this tool to a single species pair for which phased data

exist (or population data exists, so statistical phasing can be envisaged) would already be informative.

- **Answer:** Unfortunately, we 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. Because of this, and the fact that we do not require information on recent demography (our results show significant levels of gene flow stopped at the latest ~400kya), we have not implemented this analysis on any of our target species.
- Third, for the specific question of the effect of species abundance on hybridization, another limitation that arises is the effect of confounding factors. Obviously, the number of barriers that accumulated in the genome of the species through time will also affect the probability of hybridization and/or introgression. How do you account for this effect? At least this point should be better discussed along lines L212-287.
  - **Answer -** That is an interesting point. However this section is based on purely presenting the hypothesis that population size may correlate with hybridisation, considering species only whose genomes were not so different, or accumulated enough genetic barriers that were unable to hybridise. We have now made this clearer at the beginning of the section. In relation to how we can account for this effect, it is challenging, and that is why we only discuss this in relation to our hPSMC timing results and very general trends. In addition, we have now implemented a bootstrap analysis on a subset of our hPSMC pairwise comparisons as suggested by Reviewer 1 to investigate the robustness of our results across the genome. This showed that the results are likely to be conserved across the genome, and the accumulation of specific barriers/regions do not impact our results. Although there could be species-specific regions of local adaption adverse to admixture, as the gene flow we discuss was relatively ancient, these regions are likely to be very small due to many generations of recombination. Furthermore, to confidently uncover such regions, population-level genomic data from each species is necessary, which is certainly something to consider for future studies, but is beyond the scope of the current work.

#### MINOR COMMENTS

- L285-287: the discrepancy between the hPSMC analysis (which shows that there is no ongoing introgression) and the presence of fertile contemporary hybrids is at odds. This could suggest that the hPSMC analysis is not appropriate to detect recent introgression events. Could you please further comment on that?
  - **Answer:** This could be the case. We have extensively commented on the potential for this in light of our results, and state that the discrepancy between our results and recent hybrids may be due to lower migration rates than hPSMC can detect. This downfall was shown in the original manuscript by



Cahill et al. However, even with this downfall, it adds rather than subtracts from our conclusion of gene flow occurring long after the species initially diverged.

- L448: Even if it seems obvious, it could help to indicate in the legend of Figure 2 that “divergence times” are in dark colours, and the “time interval during which gene flow ceased” is in light colour. Please, add an “s” to the time interval[s].
  - **Answer:** This has been added to the legend
- Supp. Tables S3, S4: I may miss something here, but I do not understand how the # of “Mapped reads” can be higher than the # of “Raw reads”?
  - **Answer -** this should say ‘read pairs’, as opposed to just ‘reads’, and has been corrected. The reason the mapped reads can be higher than the raw read pairs is if the R1 and R2 reads do not overlap and are therefore mapped separately.

*Reviewed by Anonymous reviewer 3*

In the manuscript entitled "Speciation in the face of gene flow within the toothed whale superfamily Delphinoidea", the authors use available genomic data from 9 cetacean species to reconstruct as best they can the demographic history of the surveyed samples.

While some results seem robust, notably phylogeny and ancestral variations in effective sizes, other analyses still require further controls to convince and reinforce the current interpretations, which are at the origin of the title of the article. In particular, I did not understand how the authors rejected speciation scenarios with cycles of (geographically related) reproductive isolation / secondary contacts?

Such a test seems to me to be missing here or I inadvertently missed it. I doubt whether it is possible to test it with the current data so far, and therefore suggest in this respect to reduce the emphasis on the conclusion of "speciation in the face of gene flow". Certainly there would have been gene flow in the past, but this does not make it a support for widespread sympatric speciation in cetaceans. This gene flow may have been periodic, interspersed with periods of allopatric barrier accumulation. Although I cannot support the title put forward in the current state of the manuscript (and of the available data), the main interest I find in the article is to prepare the next step, which would be the acquisition of population data of the 9 cetacean species studied, and then to use explicit model comparison methods to statistically evaluate alternative speciation scenarios.

My overall opinion on this article is therefore positive and promises exciting future studies on this topic, studies which will provide a better answer to the question currently being asked and which need to temper a little more the conclusions that can be drawn from the current results.

**Answer:** We would like to thank this anonymous reviewer for assessing our manuscript and providing detailed comments on how we can improve our manuscript and test for the robustness of results. We have provided responses to each point and some additional explanations for some of the points that we had not made clear enough in the first submission.

**Comments :**

The title is a bit strong. I found nothing in the paper that rejects the scenario of past cycles of repeated geographic isolation followed by secondary contacts. An explicit comparison of different models would be needed to be able to test that the inferred past migrations are not past secondary contacts following older isolations.

**Answer:** The reviewer is correct that geographic isolation and secondary contact could also explain the pattern we see in our data. We originally proposed our results under the assumption that "secondary contact" would still be characterized as post-divergent gene flow under a model of parapatric speciation - as summarised with the killer whale example in the introduction. We see now that we were not so clear with some of our wording, and have slightly rephrased areas where we mentioned "continued gene flow" for more conservative language.

The introduction refers a bit too much to a vision of diversification as a series of dichotomies. It has been now accepted that gene flows between species are common and that hybrid zones of semiisolated species are frequent. Perhaps this emphasis should be reduced a little in the introduction, on a point that is certainly exciting and worth mentioning, but not so new nowadays.

**Answer:** As the aims of this study are highly intertwined with the speciation paradox in the marine realm, heavily reliant on allopatric speciation being unable to explain the diversity seen, we set up the introduction accordingly. We have presented examples where post-divergence hybridisation has been presented before and feel if we were to remove the emphasis on gene flow, then we would not be able to set up the readers not so familiar with the topic to understand the context, as well as the aims, of our study.

Figure 1:

It is possible that phylogenetic discrepancies are linked to the GC content (%GC) of the windows considered. This can be explained by the GC-Biased Gene Conversion (gBGC) known to be very strong in vertebrates. More recombination means more gBGC, and so, more elevated %GC. But more recombination also means less genetic interference (because reduced linkage) as well, and therefore less ILS.

Can the authors look at the proportion of alternative topologies for the 33% of windows that have the highest %GC, the 33% of windows that have an intermediate %GC and the 33% of windows that have the lowest %GC?

**Answer:** We have looked into whether the GC content of the window may play a role with the topology, by splitting our 50kb windows into three separate bins based on the reviewers suggestions. We have now included these results as a supplementary table S2 presenting the 5 topologies that occur most frequently for each bin, together with their counts. While there are some differences between the absolute values, the topologies are still consistent regardless of the bin, providing confidence that the GC content of the window was not the only factor driving the production of the alternative topologies.

D-foil : Same remark as for phylogenetic discordances and D-foil. Do the regions that recombine the most (at least, those that have suffered the most from gBGC) have different D-foil than the AT-rich regions?

**Answer:** This is also a valid point. However, as the phylogenetic analyses suggested that the high or low levels of GC content (and in turn low or high levels of AT content) did not cause a substantial bias, we would expect the result to hold for the Dfoil analyses. Furthermore, the fact that the D-foil analysis used 100kb windows as opposed to 50kb windows, we expect even less influence of GC content on these results due to increased data. Moreover, as Dfoil requires 5 taxon as input, the evaluation would require that we repeat this analysis 186 times (3 GC content bins x 62 5-taxon combinations), which would be very time consuming. Therefore, we opted to not look into this effect directly for the D-foil.

Cessation of gene flow : I personally did not understand in the material and method what exactly was done. This is what I understood :

Step 1) Run PSMC on a diploid genome composed of two haploids which are the consensus of 2 different species. This would apparently give the size of the common ancestor.

Step 2) Using PSMC's estimated size and predefined divergence times to identify a gene flow stop using coalescent simulations.

I don't understand what are the effective size values between the present time (at the time of sampling) and the past time (just after the split) that have been used. How simulations with ms can be used to estimate parameters. Do the authors use ABC to estimate some parameters while fixing the others?

As I am not familiar with estimating gene flow parameters with PSMC, would it be complicated to imagine a robustness check of this inference?

This could be easily done by :

i) a simulation step of the pseudo-F1 hybrid of 2 species using ms. This pseudo-F1 can be produced for two types of scenarios: without migration (Strict Isolation model) and with ancestral migration (which takes place between Tsplit and Tam where Tam is the stop time of the ancestral migration). For this, it is possible to use a few tens (hundreds) of random combinations of Ne, Tsplit, Tam considered as being the pseudo-F1 hybrids.

ii) repeat steps 1 and 2 which have been applied on the empirical data sets but on the pseudo-F1 produced, but for each of the pseudo-F1 hybrids produced using known parameters. This would make it possible to convince a little more that the practice put in place by the authors makes it possible to estimate Tsplit and Tam with precision.

**Answer:** This method is not our conception, but a previously published method by Cahill et al (<https://dx.doi.org/10.1098%2Frstb.2015.0138>). As we had performed so many comparisons (36 pairwise comparisons) we uploaded the figures to a repository, as opposed to in the supplementary material. We see that if one does not have a visual image of the output, it is hard to interpret the method and our results. For this analysis, we used all of the scripts available on github that accompany the paper (<https://github.com/jacahill/hPSMC>) which enable production of the empirical results as well as the simulations. In the original study by Cahill et al, they used simulations while specifying a set pre-divergence Ne (at the point in time just prior to an exponential increase in Ne), time of divergence, and various migration rates. They showed that although a set divergence time was simulated, the exponential increase in Ne (indicative of divergence) would shift along the time axis (or completely disappear) based on levels of migration. Based on these results the authors, who originally set out to use the pseudodiploid PSMC to test for when populations diverged, then changed to state that this method is more suitable for when significant levels of gene flow ceased between the populations.

The result shown in Figure 2-A is astonishing. While the estimated split times vary enormously, this is not the case for Tam. The cessation of gene flow seems to take place simultaneously for each pair, and this for pairs of species with different levels of divergence at the precise time of cessation of gene flow. As if molecular divergence in Delphinoidea has no impact on reproductive isolation, which is not intuitive. Is it possible to represent the shown results not in terms of years, but in terms of the expected net divergence? This could

be reconstructed for each pair by simulating the scenario inferred by the authors in order to obtain empirically what the level of divergence was at the moment when the gene flow was interrupted. This would make it possible to see whether the Delphinidae make gene flow for levels of divergence greater than 2% (to the right of the grey zone of speciation) or not.

**Answer:** We could estimate the genome divergence using the given date of the end of gene flow and the mutation rate that we have calculated. However, our mutation rates are based on the published species divergences, therefore if we would do that it would still present the same relative difference to the divergence time. This is the main reason we refrain from talking in years when interpreting our hPSMC results but rather in percentages of the branch length. We also find it interesting that gene flow appears to cease between a lot of species pairs at approximately the same time even though they should in theory have different levels of divergence. We are not really sure what may have caused this phenomenon but have some hypotheses. One explanation for this could be indirect gene flow. For example if we have three species - A, B, and C ((A,B),C) - If A and C exchanged genetic material and A and B exchanged genetic material, then we may expect the genomes of B and C to look like they have exchanged genetic material directly. This may lead to the false conclusion that B and C underwent gene flow until the same points in time that gene flow ceased between A and C (assuming gene flow will cease between A and B later). However, as this is still a speculative hypothesis, which is difficult to test with our data, we decided not to mention this as, although it is indirect, gene flow is still occurring between species B and C, which does not override our hypothesis that most Delphinoidea underwent speciation despite post-divergent gene flow events.

Interspecific hybridisation : I find the proposal of an increased interspecific introgression during periods of low abundance interesting but obviously does not seem to be exclusive. What is suggested here is that for a small  $N_e$ , selection against hybrids would be less efficient and would allow more efficient transfer of alleles between species, facilitated by drift. However, this may also be the case with large  $N_e$  when, in a sub-region of the species distribution range, the local density of individuals decreases. This process would be found again without observing any  $N_e$  related effect.

An asymmetry in  $N_e$  could also explain an increase in introgression. Notably by pump effects with species with small  $N_e$  (and therefore with many weakly deleterious mutations) which would introgress alleles without burden from a species with large  $N_e$ . While studying the  $N_e$ -introgression relationship is an interesting idea, the large number of possible and different scenarios makes highly difficult to distinguish between them.

**Answer:** We agree with the reviewer that it is difficult to distinguish between scenarios. This section was purely meant to set up hypotheses for future studies, and how different species pairs support the hypotheses. We have now mentioned that determining the direct cause of interspecific hybridisation can be difficult and a lot of factors can play a role and that low population sizes may be one of these factors.

Figure 3: are the effective sizes shown on panels B, C and D really expressed as  $10^4$  minus 4 individuals?

**Answer:** That is a typo and the minus has been removed

Conclusions : The authors insist on cases of speciation with frequent gene flow in Delphinoidea. While they show that there was past gene flow, they did not reject scenarios with past allopatry/secondary contact cycles.

**Answer:** We have taken the reviewers comment into consideration and have tried to reduce the emphasis on continued gene flow and include the possibility of secondary contacts also being an option.