Reply to Roux *et al.* "Accounting for semi-permeability to gene flow using Approximate Bayesian Computation improves inference into the history of speciation: application to a mussel hybrid zone"

1 August 2013

1 Summary

This paper addresses the important problem that effective rates of gene flow may vary across the genome and that this can interfere with demographic inference based on supposedly neutral markers. The authors claim that by allowing for genome-wide heterogeneity in geneflow, it becomes possible to differentiate alternative scenarios of population history. A hierarchical ABC approach is borrowed from an earlier publication this year (with the same first author) and applied to a pair of closely related mussel species. The authors conclude that, in their case, models allowing for heterogeneous gene flow have higher support than those with homogeneous gene flow. Among those with heterogeneous gene flow, the scenario with a long period of speration (allopatry) followed by a recent phase of secondary contact (SC), best explained the mussel data. The difference to the second model (isolation with migration and heterogeneous gene flow) seems small, but the authors give estimates of migration rates, split and contact times only for the SC model. The paper does not explicitly model the mechanistic causes for variation in gene flow (linked selection). I have concerns regarding the main focus of the paper, the methodological approach and the interpretation of the results, as well as a number of specific comments.

2 General issues

1. It was not clear to me what the main focus of paper should be; either it is to i) make the point that genome-wide heterogeneity (GWH) in gene flow may affect demographic inference, or it is to ii) show how ABC can be used in this context. If i) applies, the relevance of this paper may be hampered by the fact that, as the authors mention, other recent studies have made the point (Sousa *et al.*, 2013; Roux *et al.*, 2013). While there is a clear methodological difference between this manuscript and Sousa *et al.* (2013), my concern is that the manuscript is to some extent redundant to Roux *et al.* (2013), both with respect to the approach (hierarchical ABC) and to the main conclusion (separation of potentially hybridising species, followed by secondary contact). This would argue against ii) being the main focus. In the end, it is an editorial question whether applying an existing approach to just another pair species is relevant enough for publication in *Molecular Ecology.*
2. One of the main arguments seems to be that allowing for GWH in gene flow made it possible to discriminate alternative *demographic* models. However, the authors mention that the mechanistic explanation of GWH in gene flow is variation in the degree of permeability due to selection at isolation/speciation genes (see l.86–94 of the ms). Hence, I wonder how a signal that is produced by linked selection becomes informative about demography. Isn't there some potential for confounding effects that could only be resolved by including into the comparison some models

that explicitly incorporate linked selection? It was not clear to me what feature(s) of the data are relevant to distinguish the demographic models.

3. Performance and accuracy of ABC may depend on the model and parameter range, and on a number of choices to be made (summary statistics, rejection tolerance, distance metric). The paper lacks a proper discussion of these issues – not to mention a couple of validation steps that should by now be standard for any ABC application. 1) I missed a check of the model specification by visualisation of the prior-predictive distribution. For instance, pairwise plots of simulated summary statistics, together with the observed point, could show that the observed point is covered by the cloud of simulated points. 2) The choice of summary statitstics, in particular in the context of ABC-type model comparison (Robert *et al.*, 2011), is problematic, but not further addressed nor validated (see a number of recent publications on this topic). 3) Only one rejection tolerance is used (l.205). A number of different values should be used and the robustness of the results tested. 4) ABC with a neural network requires additional choices for tuning parameters (see l.208–209, for instance).
4. Whereas the authors have validated their ABC-type model comparison following Fagundes *et al.* (2007), they have not validated parameter inference within chosen models. This should be standard by now (pick pseudo-observed data sets from simulations, perform ABC on them and report measures of accuracy). Related to this, the authors use a conditional two-step procedure for parameter inference (locus-independent parameters first, locus-specific parameters after), which has not been validated in the context of their models and parameter schemes. There are recent studies in the ABC literature that provide theoretical guidelines and suggest empirical procedures for validation of such two-step inference.
5. The support of the SC model with heterogeneous gene flow relative to the IM model with heterogeneous gene flow does not seem substantial to me (Table 2). I would like to see estimates of migration rates (and other parameters) inferred under the IM model, too. The data might be almost equally well explained by a long period of complete separation followed by secondary contact (SC model) and a long period of continuous genetic exchange at a low rate (IM). This is not appropriately discussed, although it could completely change the conclusion about the history of the two species of mussels. Similarly, the data might also be explained by an SC model in which gene flow was interrupted or resumed at different points in time at different barrier loci, rather than by different rates of gene flow.

3 Specific comments

3.1 Abbreviations used

Q Question

C Comment

S Suggestion

R Re-formulation or change needed (usually followed by a suggestion)

→ Suggested change/correction

3.2 Title

l.1 S: Omit ‘using Approximate Bayesian Computation’ (see comment to l.24–30 below).

3.3 Abstract

l.21 R: ‘. . . between parapatric populations undergoing speciation with gene flow,. . . ’ → between populations undergoing speciation with gene flow

l.24–30 C: Here is where I first struggled in identifying the main focus of the paper (see general issue 1 above). Is it to show that ABC does a better job than other approaches in model comparison and parameter estimation in the context of variable migration rates? Or is it to make a point for the importance of accounting for variation in (effective) rates of gene flow when fitting models. I presume the second one should be the main focus (see l.103–105 in the Introduction) and I suggest rephrasing the abstract accordingly. However, as the authors acknowledge, two recent publications already looked at this problem (see general issue 1). A third option would be to focus more on the pair of mussel species and learn about their history.

l.32–33 C: I felt uneasy about this argument. The fact that when you allow for rates of gene flow to be different between loci you are able to distinguish alternative models, but when you constrain on homogeneous gene flow you are not, is not *per se* evidence in favour of heterogeneous gene flow.

3.4 Introduction

l.38 R: Delete 'recent'. **R:** Delete 'by evolutionary biologists'

l.40 S: Either refer to review articles exclusively, or give a more representative list of references.

l.40–42 Q: Isn't the main challenge that, as models become more complex, parameters and models may be confounded (the data may not be informative for resolving such conflicts)? Actually, models do not need to be very complex for this to happen.

l.43–44 C: 'parapatric populations' and 'speciation with gene flow' reads redundant. See same comment to l.21 of the Abstract

l.52–53 R: '. . . heterogeneity in genomic patterns of differentiation. . . ' → . . . heterogeneity in genetic differentiation. . . .

l.58–64 C: This part fails to distinguish between models and methods/approaches. I would classify IM as a model in this context (that comes with a set of assumptions), not a method; various approaches (explicitly likelihood-based ones, ABC,. . .) can be used to do inference under the IM model. You are concerned about the limiting assumptions, for instance, of homogeneous gene flow. (Unfortunately, the software is called IM, too. This causes the confusion).

l.68 R: 'method' → model

l.76 R: 'For example, the best supported scenario. . . ' → For example, in the study by Roux et al. (2013), the best-supported scenario. . . .

l.76–79 S: The formulation is not clear. How can one scenario (model?) be supported by another model? The data can support one or the other model in the sense that one model is more likely compared to the other one, given the data. Please clarify.

l.102 S: '. . . for which the IM method. . . ' → for which inference under the IM model. . .

l.107–109 Q: What explicitly is new compared to Roux *et al.* (2013) with respect to the method/approach?

3.5 Materials and Methods

l.128–130 C: This sentence was not clear to me. What is meant by a 'single allele per individual'?

l.141 R: '. . . departure of site frequency spectrum from. . . ' → . . . departure of the site frequency spectrum from. . . [or: . . . departure of site frequency spectra from. . .]

l.142 S: Insert comma after '(Tajima 1989)' and 'using. . . '

l.170 S: No hyphen in 'scaled-Beta' (?)

l.173 S: It might be good to remind the reader of what the mean and variance of the beta distribution are in terms of α and β. You could even give two numerical examples of combinations of α and β, for instance, one leading to a U-shaped distribution of migration rates across loci, and the other one to a bell- or L-shaped one.

l.174 R: 'large' → wide

l.175 S: I suggest placing the definitions of the various θs first.

l.174–190 C: Please justify the choice of the priors. **S:** Use a standard way of specifying an interval, *e.g.* $[a, b]$ instead of a-b.

l.181 C: It is a matter of taste, but I suggest not to start a sentence with a symbol or abbreviation (this applies to other places in the ms, too).

l.182–183 R: 'We sampled $T_{\text{split}}/4N_{\text{ref}}$' from the interval of 0–25 generations, 0–10^7 generations in demographic units' → We sampled $T_{\text{split}}/(4N_{\text{ref}})$ from the interval $[0, 25]$, which corresponds to $[0, 10^7]$ generations.

l.187 Q: Is the interval of 0-30 correct here? Compare to 0-20 in l.185.

l.187–190 Q: What was assumed about the scalar c (see l.171)?

l.198–202 Q: How is this two-step procedure for model choice justified? I can see the practical reason for doing so – avoiding too many models to be compared at once, which would force you to relax the rejection tolerance (n). But it could be that the homogeneous and heterogeneous version of a given model, say the IM, both explain the data far better than the homogeneous and heterogeneous version of, say, the AM model. In that case, your two-step procedure would keep one of the AM models and throw out one of the IM models. However, it would seem justified to keep both IM models and omit both AM models. It should be straightforward to empirically check if a deviation from the two-step procedure changes the main conclusion (the ranking of the various models). PS: Your results suggest that this is not an issue (l.260–262); perhaps state this early on so that the reader does not worry.

l.205 C: The rejection tolerance is known to have an impact on ABC inference (Beaumont, 2010). I would therefore like to see a comparison for a (reasonable) set of values.

l.216–218 C: It is a matter of taste whether one wants to bring up p-values in a Bayesian setting.

l.221–222 R: I see the practical advantage of this two-step estimation procedure, but it is by no means theoretically valid in general. It needs to be justified for each model separately, which involves showing at least empirically that parameter estimates obtained this way are not significantly worse than those obtained by joint inference (for a detailed treatment of conditional ABC inference, see Bazin *et al.*, 2010; Aeschbacher *et al.*, 2013).

l.223 R: Please say why you transformed the parameters. Was it to avoid projection of posterior density out of the prior range? Was this an issue at all?

l.230 R: 'Hence,. . . ' → Specifically,. . .

l.232 C: Again, this procedure is not valid in general (see comment to l.221–222 above), but needs to be justified. It depends on (approximate) sufficiency and ancillarity of summary statistics used for the different sets of parameters.

3.6 Results

As a general point, I am worried that the observed data might also be explained by differential onset of reduced gene flow at various speciation/isolation loci, rather than heterogeneity in rates of gene flow among loci. All of the models considered here assume that the temporal parameters are identical among loci.

l.239 C: I stumbled over 'mulitply captured', as it might mean several things. Please make sure it is properly described in the Methods, and then it can be omitted here.

l.249–252 Q: Although intuitively clear, is there a reference you could cite?

l.291 S: Omit 'and that we explored the correct parameter space'. See point 1) in general issue 3 above.

l.292 R: Please introduce the abbreviation ‘HPD95’ here.

l.295 R: Remove ‘, but not significantly,’, as it is not clear what ‘significant’ means here.

l.298 C: The formulation is unfortunate, because a scenario as such cannot be informative. Information is in the data, and a model/scenario can be more or less likely given the information in the data.

l.306 R: ‘. . . the estimated joint shape parameters. . . ’ → the jointly estimated shape paremters (?)

l.309 C: The goodness-of-fit procedure is not described and potential references are missing. **R:** ‘joint-posterior’ → joint posterior

l.308–312 C: This is only a partial check. Other models might have been misspecified (see point 1) in general issue 3 above). **Q:** How can the underestimation of F_{ST} be explained?

l.311 R: Delete ‘by the scenario’.

l.314 C: I suggest avoiding the phrase of an ‘informative (posterior) distribution’ without further specification. A flat posterior is also informative in a sense.

3.7 Discussion

Parts of the discussion are redundant (first and last paragraph), whereas a series of methodological issues and interpretational uncertainties are not appropriately discussed (see detailed comments above). I suggest shortening the existing content and adding a subsection on these limitations.

l.341–344 C: Strictly speaking, it is not clear what is meant by ‘efficiency’. Efficiency in terms of computation time at the cost of strongly reduced accuracy would not necessarily be desirable. In that sense, efficiency that does not take into account accuracy is not a good criterion. Please clarify what is meant here.

l.354 C: A former comment seems pending here.

l.362–363 C: The sentence ‘It may also miss information provided by the model itself.’ is not clear to me. Do you mean that the model might make invalid/restrictive assumptions?

l.375–377 C: This sentence could give the wrong impression that using a hierarchical design is the only way of obtaining the distribution of migration rates across loci. However, it is a compromise between specifying one prior for all loci (which does not provide the desired distribution) and specifying individual priors for each locus (which would of course provide the desired distribution).

l.382–383 C: ‘loosely permeable’ was not clear to me. I think you mean relatively resistant to introgression (or similar).

l.398 R: Delete the second ‘both’.

l.400 R: ‘. . . will now be confirm this. . . ’ → . . . will now be to confirm this. . .

l.406 S: Insert a comma after ‘among loci’.

l.426–427 Q: If neglecting GWH in gene flow can lead to statistical support for an incorrect model, can enforcing GWH in gene flow not do so as well? Does it not depend on the broader context of assumptions? What criterion is used to do a ‘fair’ comparison, taking into account the additional degrees of freedom for more complex models (which tend to increase likelihoods)?

4 References

The formatting of the references does not come up to the specifications of *Molecular Ecology*.

l.457 R: Give full journal name.

l.459 R: Give full journal name.

l.463–464 R: Change species names to italic. Give correct journal name

l.466 R: Give correct journal name.

l.468 R: Change species names to italic. Give full journal name.

l.470 R: Change species names to italic.

l.477 R: Change species names to italic. Give full journal name.

l.479 R: Give correct journal name.

l.482 R: Give full journal name.

l.491 R: Give full journal name.

l.496 R: Change first letters to upper-case in journal name.

l.502 R: Change species name to italic.

l.505 R: Change first letters to upper-case in journal name.

l.5012 R: Change first letters to upper-case in journal name.

l.517 R: Give full journal name.

l.531 S: Please check journal name.

l.542 R: Give full journal name.

l.543–544 R: Change title to lower-case.

l.545 R: Give full journal name.

l.557 C: Journal name, issue number and page range are missing.

l.560 R: Please check journal name.

l.565 R: Give full journal name.

l.569 R: Change first letters to upper-case in journal name.

l.577 R: Please check name of publisher and add address/city.

l.580 R: Give full journal name.

4.1 Figures and tables

l.602 R: ‘nascent’ → derived

l.604–605 R: Please make this clearer; N and m are not the same for M_1 and M_2, are they? Use another formulation than ‘are expressed in x units’.

l.607ff. Q: Further caption missing? Explain the x-axis (including a reference to Kimura?)

l.609–610 S: I suggest using ‘homogeneous’ and ‘heterogeneous’ instead of ‘homo’ and ‘hetero’. This applies to several places later in the figure and table captions.

l.661 R: ‘best-support’ → best-supported

l.622–623 S: ‘…for the two homo and hetero alternative SC models.’ → …for the homogeneous and heterogeneous version of the SC model.

l.629–630 C: This was not clear to me.

Table 1 R: 3rd line of caption: ‘silent’ → Silent

Table 2 R: Caption: ‘mode’ → model; Add a fullstop at the very end. **Q:** Is the differences between the the posterior probabilities for the AM/homogeneous (0.3206) and AM/heterogeneous (0.6794) substantial enough for omission of AM/homogeneous? Similarly, is the difference between IM/heterogeneous (0.4016) and SC/heterogeneous (0.5194) substantial (see general issue 5) above)?

Table 3 S: Please remind the reader of the unit of the time parameters. I think it is generations here, but in the literature, uppercase letters are often used to describe time on the scale of $2N$ generations.

Figure 1 R: Please make clear that in the SC model, migration is also allowed to be asymmetric

(add M_1 and M_2 to the bottom-right panel). That is at least how I understood it.

Figures 4–7 R: Please add labels to the y-axes.

5 Supporting Information

5.1 Supporting figures

Suppl. Fig. 1 R: Please use unambiguous formal expressions for the conditional probabilities and define them properly. Avoid 'model-A' (the hyphen). The label of the x-axis is not consistent with the expressions used in the caption.

Suppl. Fig. 2 Q: Should the first line be bold face? l.650: '. . . of the four models at the SC posterior probability = 0.5194. . . ' → . . . of the four models at an SC posterior probability of 0.5194. . . . **Q:** Why was the cut-off chosen to be 0.5194 exactly? **C:** The y-axis has no label.

5.2 Supporting table

Suppl. Tab. 1 R: Please unify the style of citation according to the citation style of *Molecular Ecology* (volume and issue, or volume only; spaces; full or abbreviated journal names)

Suppl. Table 4 and 5 R: Please define in the caption what exactly is meant by 2.5% and 95%. Are these the limits of the 95% HPD interval?

References

Aeschbacher, S., A. Futschik, and M. A. Beaumont, 2013 Approximate bayesian computation for modular inference problems with many parameters: the example of migration rates. Mol. Ecol. **22**: 987–1002.

Bazin, E., K. J. Dawson, and M. A. Beaumont, 2010 Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. Genetics **185**: 587–602.

Beaumont, M. A., 2010 Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. **41**: 379–406.

Fagundes, N. J. R., N. Ray, M. Beaumont, S. Neuenschwander, F. M. Salzano, S. L. Bonatto, and L. Excoffier, 2007 Statistical evaluation of alternative models of human evolution. Proc. Natl. Acad. Sci. U.S.A. **104**: 17614–17619.

Robert, C. P., J.-M. Cornuet, J.-M. Marin, and N. S. Pillai, 2011 Lack of confidence in approximate bayesian computation model choice. Proc. Natl. Acad. Sci. U.S.A. **108**: 15112–15117.

Roux, C., G. Tsagkogeorga, N. Bierne, and N. Galtier, 2013 Crossing the species barrier: Genomic hotspots of introgression between two highly divergent ciona intestinalis species. Mol. Biol. Evol. **30**: 1574–1587.

Sousa, V. C., M. Carneiro, N. Ferrand, and J. Hey, 2013 Identifying loci under selection against gene flow in isolation-with-migration models. Genetics **194**: 211–233.