

*Associate Editor's comments to author:*

*Associate Editor: Drummond, Alexei*

*Comments to the Author:*

*Your revised manuscript has been reviewed by one of the original reviewers and a new reviewer. As you can see the original reviewer now finds the article acceptable. Reviewer 2 has made a diligent effort to check several points and has made comments that need to be addressed in the final submission. The second reviewer has also carefully checked the revised document against comments from the first round of reviews and again finds some points that require clarification, explanation or minor changes to the text. In revising this manuscript please ensure that you address all of Reviewer 2's comments. I suspect that you should be able to do this without major changes to the manuscript and I hope that I will be able to accept without further review your revision should you carefully address these comments.*

Thank you. I thank Reviewer 2 for their close attention to detail and help in further clarifying the manuscript.

*Reviewer(s)' comments to author:*

*Reviewer: 1*

*Comments to the Author*

*The concerns I had with this manuscript have been addressed, and the more significant issue I had with regard to rate estimates associated with Mike Hickerson's data points has been clarified. I'm not familiar with these species and the differences in generation times that exist among them, but I now follow the logic applied by the authors. It might be worthwhile adding an explicit statement somewhere early on in the manuscript that all rates are expressed in terms of generation time, as it isn't clear that the fossil calibrated rates have been corrected in the same way (I'm assuming they have been, but it would add an extra level of assurance to the reader if this was stated).*

I'm glad I was able to clarify anything that may have led to misunderstanding. I have kept all rates throughout the manuscript in units of per-lineage per million years to facilitate comparison with other published rates. As requested, I have now added a statement in the methods section P9 L1-2: "Unless stated otherwise, we report all rates in this paper as lineage mutation rates (= 1/2 divergence rate) in units of percent change per million years (%/my)." For one section of the discussion and figure 4, I have clarified that I am using percent change per million generations.

*This manuscript makes a compelling argument for elevated rates among the focal species, and for elevated rates over more recent time intervals. I am however sceptical as to whether this is the real explanation behind these patterns. We know that population growth leaves signatures in sequence data that can be extracted with methods such as those applied by the authors. Something that has been less investigated is the extent to which alternative demographic scenarios may present similar mutational signatures. The kind of scenario I am thinking about is one of small (refugial) populations, with limited genetic connectivity among themselves, that initially share a common allele, or set of alleles. In the context of the Sunda shelf study system, and rising sea levels, these populations would subsequently coalesce as the Sunda shelf became inundated. The question that nags me is whether a population history such as this could have generated the observed sequence patterns. While there may be no signatures of population structure now, the question that I think is interesting is to what extent might there have been structuring in the past, something that violates the panmictic population model, and what would the implications of this be for contemporary patterns of genetic variation. The authors mention there to have been a history of refugial populations prior to the inundation of the Sunda shelf, which covers a period of approx 120,000 years. I'd suggested in my previous review that I'd like to see the authors do something to address this, but perhaps this is beyond the scope of the manuscript. However, I think this would be a more compelling paper if some form of simulation analysis could be applied to assess this.*

I agree with reviewer 2 that simulations of complex scenarios such as this are beyond the scope of this paper. However it seems to me that if current gene flow has erased previous genetic structure between refugial populations to the extent that it is not detectable by *Fst* estimators, then what we have is a single, larger population. This is best modeled as a population expansion.

*Brent Emerson*

Reviewer: 2

Comments to the Author

*This is a nice study, showing a novel approach and providing further support for the dependence on calibration times of rates of evolution. The authors present an approach that uses non-genetic information to condition inference from genetic data; such approaches will become more important, as models are more complex and need some justification and calibration. I agree with the previous reviewer 2 that this paper is of importance to the broader field. I have a couple of conceptual issues -- four of them major --, and some editorial suggestions. Later, I refer to the concerns raised by two previous reviewers and the author's replies thereupon.*

Major issues

=====

a)

*In the footnote to table 2, the authors give the formula according to which the lineage mutation rate  $\mu$  [mutations per year per site per million years] was obtained as  $\mu = (\tau/b)/(2c)$ , where  $\tau$  is time measured in mutations per sequence [mutations/sequence],  $b$  is the number of sites per sequence [sites/sequence],  $c$  is one of the two calibration times [years]. First, there is a bracket missing -- it should read  $\mu = (\tau/b)/(2c)$ . The units work out correctly: [mutationn/sequence]\*1/[sites/sequence]\*1/[years] = [mutations/(sites\*year)]. However, numerically, I was not able to reproduce the lineage mutation rates  $\mu$  given in table 2 for the mismatch distribution approach. For instance, with *T. crocea* and for  $t = 19,600$  years, we have  $\tau/2 = 1.2 \cdot 10^{-2}$  mutations/sequence, and  $b = 485$  sites/sequence (according to table 1). Multiplying by  $10^6$  to obtain the rate per million years, this results in  $1.2 \cdot 10^{-2} \cdot 1/485 \cdot 1/19,600 \cdot 10^6 = 0.00126 = 0.126\%$  [mutations per site per million years]. This is not the same as 61.14% as given in table 2. Some clarification is needed.*

Thank you for catching this. The number given in the third column, labeled  $\tau/2$ , is actually  $\tau/2b$ , where  $b$  is the number of sites. I provided this number because it is directly comparable to the value for  $t_{\text{transition}}$  given for the BEAST model below it in the table. I have amended the label for this column, as well as the formula in the footnote.

b)

*The authors first infer Bayesian skyline plots (BSP) from genetic data; those provide an estimated trajectory of  $N_e$  across time. This trajectory is then used to devise an appropriate model of demography. The authors find that a two-epoch model -- a period of constant  $N_{e1}$  followed by logistic growth -- is appropriate. This model is then employed, together with coalescent theory, to establish the mutational depth (i.e. the time in number of mutations) at which population growth started (transition time  $t_{\text{transition}}$ ). It is not clear to me, if the BSP analysis assumes constant or variable mutation rate  $u$  across time. If it does, isn't there a circularity in the procedure of inference? One first uses genetic data to infer a trajectory of  $N_e$ . Then one uses the same data to infer mutational depths. On the other hand, if constant substitution rates were not assumed in the BSP analysis (which, as I understand Drummond et al. 2005, is feasible), why do the authors not infer the substitution rates directly and jointly with the trajectory of  $N_e$ ? Some more clarification would be helpful.*

Our BSP analysis assumed a strict clock model, however the above description of our inference procedure is not quite correct. Inference of  $N_e$  in the BSP framework is made from the number of coalescent events across multiple intervals each of which is a parameter in the analysis. Because the sizes of these intervals are estimated during the analysis, the BSP cannot include a parameter for time of expansion, but can provide an overall schematic for the history of  $N_e$ . We therefore used the BSPs to provide priors to the 2-epoch analysis, which is a simpler model that can get attracted to numerous alternate histories, but does include a parameter for time of expansion. We are therefore "fitting" (in a Bayesian context) this 2-epoch model to the history inferred by the BSP. This is no more circular than iterative procedures used to fit a regression model to data.

It might be possible to simultaneously infer mutation rate if  $t_{\text{transition}}$  were also given strong priors, but I view this as the subject of future work.

I have clarified that the BSP uses a strict clock in the methods section. I feel that I've already discussed strengths and weaknesses of our inference procedure under *Sources of Error*: "By setting priors of the two-epoch model to confidence intervals for the Bayesian skyline plots, we essentially estimated expansion times that fit the BSPs ... However, it is important to note that our estimates will also be affected by any violations of the

assumptions of the Bayesian skyline method”

c)

*The effect of rate variation across sites is not mentioned, although this could have an impact on the inference conducted in this paper (e.g. Arbogast et al. 2002). At least in the Discussion, this should be mentioned.*

None of the molecular evolution models inferred from ModelTest included among-site rate variation as a parameter. This is not surprising given the short timescales involved, but this is where this variation would be accounted for in our model. I have changed P18 L7 in the discussion to reflect this.

d)

*Fu's  $F_s$  statistic (table 1), a bimodal observed mismatch distribution (figure 2) as well as the model comparison in table 2 suggest that the model of constant size cannot be rejected for *T. crocea*. E.g. on page 13, lines 1-2, it says that "none of the datasets rejected a sudden expansion model". However, it should also be mentioned here that the constant-size model could not be rejected for *T. crocea*. The point is made later in the Results, but, to my impression, not emphasized enough. In the Discussion, it is mentioned appropriately (page 17, lines 8-10).*

The sentence in Results P13 L1-2 (Now L9) is specifically referring to results from the mismatch distribution, and should therefore remain unchanged. As stated later in results, the two-epoch model for *T. crocea* was actually better than a constant-size model, but only marginally so. *T. crocea* has therefore been removed from further inference in the discussion and also in figure 4.

*Minor issues*

=====

a)

*Avoid jargon in the Abstract, such as "mutational depth" (use e.g. time measured in number of mutations?) or "mismatch distribution" (use distribution of the number of pairwise distributions?). Then, early in the introduction, define these terms (e.g. on page 2, line 15, or on page 6, line 17).*

I've changed "mutational depth" to "time in mutational units". I believe that the mismatch distribution is well known to readers of MBE and attempts to generalize it would be excessively wordy. I've also added a citation to Rogers and Harpending 1992 to the first place where the mismatch distribution is mentioned.

b)

*Early on in the Introduction, a short listing of potential reasons for rate variation as a function of calibration time would be helpful. Several reasons are discussed throughout the paper, but a concise description early on would help (e.g. distinction between substitution and mutation, change in  $N_e$ ,  $s$ , actual changes in  $\mu$  (per site mutation rate), etc.).*

I feel that potential reasons for observed time-dependency are briefly and appropriately reviewed from P4 L20 to P5 L3. These include sequencing errors, calibration errors, saturation and purifying selection. More thorough discussion of these factors should be saved for a review article.

c)

*Page 4, lines 2-4: If population divergence is overestimated, does this not lead to \*underestimation\* of the molecular substitution rate? At least, I could not find the direct conclusion as stated by the authors in the reference they give (Arbogast et al. 2002).*

Measuring the divergence depth of the gene rather than of the population will lead to a higher value for divergence depth in mutational units. This will translate into an overestimate of the rate when calibrated by the same geological time of divergence. I've added a parenthetical to clarify that the time referred to is in mutational units.

d)

Page 4, lines 20-21: To avoid misunderstandings, I would rephrase the sentence to "In addition to avoiding -- through coalescent models -- the problem of ancestral polymorphism, their analysis suggests that...". Otherwise, it is not clear whether the lineage sorting issue is *\*due\** to the coalescent (which it is not) OR whether the coalescent helps to account for it (which it does).

Changed.

e)

Page 5, lines 21-23: Perhaps avoid using "use" twice in the same sentence? E.g. "We demonstrate the method for a recent....". OR: "We apply the method to a recent..."

Changed.

f)

Page 6, lines 19-23: Reformulate that sentence; it is too long and potentially confusing. Yet, it contains the main idea of your approach! So, it should really be well understood.

Thank you. I split the sentence into two at the "which" clause.

g)

Page 9, lines 1-7: Here, you address concern 3 by reviewer 1 -- which is fine. However, shouldn't you mention that your estimate is an approximation, since you do not account for the projection error (the slope of the shelf)? Maybe your ArcMap tool does it? It would be good to comment on it, at least. If you have not accounted for the slope, then I would presume that the true surface that became available to marine species was larger. The dashed curve in figure 1 would then be steeper, I suppose.

True enough. I've added a sentence to this effect.

h)

Page 9, line 23: To avoid misunderstandings, write  $\mu = \tau/(2c)$  instead of  $\mu = \tau/2c$ .

Changed.

i)

Page 10, lines 1-2: Make the scaling *\*explicit\**, i.e. say that you *\*divide\** by  $b$  (see footnote to table 2).

Changed.

j)

Page 10, line 3: Perhaps avoid the term "demographic history information"? E.g. "To assess the information available in each genetic dataset on demographic history, ..."

Changed.

k)

Page 10, line 12: There is no *\*uninformative\** prior; even a uniform or Jeffery prior imply information, e.g. due to the choice of the boundaries and/or the scale. I would avoid using "uniform"; just state what you choose and, if possible, justify.

I've changed it to "relatively" uninformative

l)

Page 12, line 7: Perhaps remind the reader of the definition of LGM.

Done.

m)

Page 12, line 15: Shouldn't it read "ModelTest" instead of "Modeltest"?

Changed.

o)

Page 12, line 21: A bimodal mismatch distribution is expected under constant size; you should state this here, and give a reference (e.g. to Hein et al. 2005, chapter 4).

Done.

p)

Page 13, lines 1-2: BUT: The constant-size model could also not be rejected for *T. crocea*! (see Major issue d) above)

The results given here are strictly for the sudden-expansion model of the mismatch distribution. Because this approach is more or less frequentist in nature it does not simultaneously evaluate a constant and expansion model, but only gives a probability of rejecting the null model, which is sudden expansion in this case.

q)

Page 16, line 6: *\*conservative\** w.r.t. what? To the hypothesis that rates of evolution increase with more recent calibration times. You could make this clearer.

Clarified.

r)

Page 39, line 29: No starred values present in figure 2?

Correct. This has been removed.

s)

Page 40, lines 4-6: Lower prior boundaries for  $t_{\text{transition}}$  and  $t_{50}$  missing? Or not visible? You should comment on it.

The lower boundary for  $t_{\text{transition}}$  was given in the text with a justification. I have now amended the text to include  $t_{50}$  and mention in the figure legend that these values are not visible in the figure.

t)

Page 44, label to y-axis in figure 4: Shouldn't it read "million years" or "million generations"? See reply to 2nd major concern by reviewer 1.

This was a misstatement in my response letter. The axis correctly reads "changes/site/million generations" in figure 4.

Issues raised by previous reviewer 1

=====

1)

Reviewer 1 doubted whether the article would take the debate about time dependency of substitution rates any further. I agree with the authors that this is a matter of editorial opinion. It is true, however, that the authors not only present a novel approach, but they do present support in favour of time-dependent rates of evolution (figure 4), accounting for uncertainty (error bars).

Thank you. It is gratifying to read this.

2)

The misunderstanding related to *\*divergence rates\** versus *\*per lineage rates\** have been clarified by the authors. However, the authors claim to have corrected the label of the y-axis in figure 4, which I could not verify (see minor concern t) above). In their reply to the 2nd major concern by reviewer 1, the authors say they have changed the label of the y-axis to "changes/site/million *\*years\**". However, the actual label now reads "changes/site/million *\*generations\**". What is correct? I suspect "generations", but the authors should clarify.

Reviewer 1 also argued that the authors had cherry-picked previous results to support their evidence. The authors have convincingly explained that this was not the case -- their choice is conservative w.r.t. to the suggestion they make. The 'minor changes' made to figure 4 to correct for new insight on generation time of the alpheid snapping-shrimp (points C and D) are not visible to me; I cannot judge if the changes have been made.

As noted above, "generations" is correct. This was a misstatement in our response letter.

3)

*Reviewer 1 asked the authors to better convince the reader that the genetic variation they were sampling postdates the low-level sea stand of the LGM. The authors explain that their approach accounts for variation before the expansion. Reviewer 1 suggests to perform ABC simulations to test more complex demographic scenarios. I do not think this is necessary, unless clear-cut alternative hypotheses can be justified.*

*The authors have added a better description of the habitat ranges that became available to the organisms studied as a function of time (but see minor issue g) above).*

Thank you. While we certainly believe that ABC simulation of more complex model is an important avenue of future research, we don't believe that it falls within the scope of this paper.

*The minor points raised by reviewer 1 have been addressed, except for the contradicting replies by the authors to the 3rd major issue versus the 7th minor issue raised by reviewer 1. (\*generations\* versus \*years\* in y-axis of figure 4). Whether or not to include the supplementary figure is a matter of editorial opinion.*

Agreed. We believe that the supplemental figure may be helpful to include for a visual orientation to the genealogies.

*Issues raised by previous reviewer 2*

=====

1)

*Reviewer 2 raised the concern that much of the signal in the data could have been caused by purifying selection. There are potentially confounding effects of selection and demography. I agree to this concern.*

*The authors have included analyses and results on further, simpler demographic models, and shown that these have less support than the two-stage models originally proposed. The idea is that purifying selection would lead to a more gradual profile of "growth" (effective change in  $N_e$ ), and if such simpler models can be rejected, purifying selection seems less likely. This addresses the concern to some degree, but to fully account for purifying selection, that process needs to be explicitly included into the models. That, however, makes (coalescent) simulations and analyses much more demanding; tools for inference/simulation are only just becoming available. Given that the authors are aware of the problem and discuss it at length, I think that the issue should not be a reason why not to publish the article.*

Thank you. We agree that this issue needs to be addressed in future research as methods such as those in O'Fallon et al. (2010) become available.

2)

*Reviewer 2 asked for more details about the habitat and distribution of the organisms studied. This is essential, because changes in sea level and the topology of the shelf are used to infer the demographic past of the species. The authors do now provide more details (see comments above and the GIS analysis described on page 9), supporting their approach.*

Thanks.

3)

*Reviewer 2 asked if there are alternative explanations for the demographic expansion than the the flooding of the Sunda Shelf. The authors admit that there are other potential reasons, but that they are confident that the rise of sea level was the most important one. I think this is fine, but the alternative explanations should at least be mentioned in the Discussion.*

We now mention some alternatives on P17 L14-15.

4)

Reviewer 2 asked if earlier marine transgressions than the one related to the LGM could have generated the proposed population expansion. The authors refer to Tager et al. 2010, who show that sea levels stayed between -100 and -20 meters below current levels between 144 kya and the LGM (18 kya). They make the point that, even if earlier geological events caused a population expansion, one would need to explain why the more recent geological changes did not leave a signal. The authors provide enough information to support their approach and assumptions. Thanks.

5)

Reviewer 2 correctly states that the mismatch distribution was originally given in terms of mutations per lineage per \*generation\*, not \*year\* (Rogers and Harpending 1992). Therefore, in table 2, the corresponding rates should be corrected by generation time. Nevertheless, the authors decided to keep the rates per \*year\* for better comparison to the two-epoch rates shown below in table 2. Since the authors state this in the main text and in the footnote, it is a matter of editorial opinion. However, it is confusing that, while the rates in table 2 are in terms of \*years\*, those in figure 4 are in terms of \*generations\* (at least as claimed by the authors in their reply to a minor concern by reviewer 1; see above). I suggest that the same unit per time is used throughout the paper; for consistency with Rogers and Harpending 1992, I would use per \*generations\*. The authors need to clarify units and calculations in table 2 and in figure 4 anyway.

I agree that it is potentially confusing to give rates in per-year and per-generation units. However, I feel it is most important that the rates throughout the paper be per million *years* since this is the unit most often used for molecular clock estimates, and facilitates comparison. However, the disparity in rates among taxa is really best explained by differences in generation time, and thus it is best, just for figure 4, to give these units. I have taken care to emphasize this in the discussion and in the figure and its legend. “Per generation” is now italicized in the legend for figure 4.

6)

No further comments.

7)

Since I could not reconstruct the substitution rates obtained for the mismatch distribution in table 2 (see my major concern a) above), I am not able to further comment on this issue.

*Additional comments to minor suggestions by reviewer 2:*

- Re-formulate the sentence on page 4, lines 20-23 (see minor issue d) above).  
Changed.

- A “ $1/x$  prior” is a Jeffrey prior. Why not state this?

We refer to this prior as a  $1/x$  prior due to comments by Marc Suchard on the BEAST Google Group (April 8, 2008) in which he states that this is only a Jeffery’s prior when the model contains only a single scale parameter.

- I also don't understand the second part of the question by reviewer 2 in the 6th comment  
Thanks. Neither did I.

I have no further comments to issues raised by reviewer 2.

Thank you for your diligent attention to detail and affirmation of our research.