

Dear Andrew, Thomas et al.,

I've taken a little time to get back to you on your revised preprint; I was glad to see your use of the reviews to improve the paper but couldn't quite parse the response to the comments regarding the spatial diffusion analyses. The original reviewer, Miguel Navascués, took an immediate further look and has clarified the point in some detail. The bottom line is that the approach is based on the same kinds of assumptions as its discrete state predecessor (in particular with regard random sampling and in ignoring population structure when calculating the probability of the coalescent tree), and despite its popularity might deliver similarly inaccurate results when those assumptions are violated. My impression is that you sampled in order to best represent the distribution, not to represent populations in proportion to their size, so this at the least does seem potentially problematic. He suggests either to remove the analysis or to include a thorough discussion of its potential problems (in the context of your data, I would add), either of which solutions should be straightforward for you to implement.

I have included some minor further suggestions in the tracked-changes version of the text which I will forward on separately as it seems the upload function here only accepts pdf. I'll look forward to seeing the revised – and doubtless final – version in due course.

All the best,  
Mike

Thank you for your advice and suggestions regarding our manuscript. We have put into place those tracked changes you suggested, leaving only the track changes for one sentence you commented on that we have clarified.

We have also added a paragraph on the issues raised by the reviewer related to the spatial diffusion method (line542), following our inquiries with the senior author of the method, Philippe Lemey.

Additional requirements of the managing board:

As indicated in the 'How does it work?' section and in the code of conduct, please make sure that:

-Data are available to readers, either in the text or through an open data repository such as Zenodo (free), Dryad (to pay) or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data.

-Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) are available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused.

-Details on experimental procedures are available to readers in the text or as appendices.

-Authors have no financial conflict of interest relating to the article. The article must contain a Conflict of interest disclosure paragraph before the reference section containing this sentence: The authors of this preprint declare that they have no financial conflict of interest with the content of this article. If appropriate, this disclosure may be completed by a sentence indicating that some of the authors are PCI recommenders: XXX is one of the PCI XXX recommenders.

## Reviews

Helmstetter and coauthors have addressed most of the comments raised in the previous round of review satisfactorily. However, my main concern has been dismissed by the authors without enough justification. In my previous review I argued that the method used to study spatial diffusion (i.e. BEAST + SPREAD3) is based in an artificial model that has not been properly validated. I recommended to remove it from their work. Authors have decided to maintain it and provide no evidence-based argument on the validity of the method to justify their decision. My position on this has not changed, these are my reasons:

In 2009, Lemey et al. (2009 doi:10.1371/journal.pcbi.1000520) presented a new method to make phylogeographic inferences. This method, often referred to as “mugration” or “discrete trait analysis” (DTA), is based on modeling spatial location as a discrete trait that evolves through a phylogeny/genealogy; that is, modeling migration as if it was mutation. This is not a process-driven model because it removes the influence of migration on the shape of the tree topology. In real life, the dynamics of migration are different to the dynamics of mutation. This is an utilitarian model. There is nothing wrong with an utilitarian model, as long as it is useful. Many of us welcomed the new method as promising, despite the fact that the article presenting it did not have any formal validation.

In 2015, after more than 500 citations of Lemey et al. (2009), most of them applications of the method, De Maio et al. (2015, doi:10.1371/journal.pgen.1005421) presented an evaluation of this method. This work shows that DTA suffers from severe biases in the estimation of dispersal rates, poor accuracy of the estimation spatial location of ancestral nodes and misleading measures of the uncertainty of the results. The authors of the method write about it:

“Despite their popularity, DTA make a number of restrictive assumptions that can be inappropriate when applied to the migration of lineages between geographic locations. DTA potentially under-represent ancestral trait uncertainty and are known to be sensitive to biased sampling of subpopulations.” (Baele et al. 2018, doi:10.1016/j.coviro.2018.08.009).

Today, Lemey et al. (2009) accumulates more than 1000 citations. Despite the evidence that it is unreliable, it stays in the phylogeographic toolbox. Many researchers learn about the methods they use on empirical papers dealing with similar questions. A single methodological article showing the poor performance of one method can easily be missed among hundreds of articles that apply the method without questioning its validity. It is therefore important that the community gains awareness of the problems that some methods have and that those problems are reflected on what we publish. At the bare minimum, acknowledgment of the limitations/problems of the methods must be presented and discussed, to warn the reader about the uncertainty of the results.

Helmstetter and coauthors argue, however, that they are using “continuous spatial diffusion” (Lemey et al. 2010, doi:10.1093/molbev/msq067) and not DTA. The difference of between them stems mainly on considering space as a continuous variable instead of a discrete variable. The core of the approach remains the same, treating space as a trait that evolves along the phylogeny/genealogy. On contrast with Lemey et al. (2009), Lemey et al. (2010) presents a validation of the method by means of simulations. However, those simulations were on the inferential model, that is, they simulated the evolution of a continuous trait on a given phylogeny and they called it “space”. This give us little information on the performance of the method on more realistic dynamics, where migration is explicitly modeled and changes both the “spatial state” of the lineages and the topology of the genealogies (such as the simulations by De Maio et al. 2015). As noted by De Maio et al. (2015), the problem of DTA is its use as a model of migration and not as a model of evolution of traits, purpose for which it was originally developed. Therefore, I believe there is reason to expect similar problems for the “continuous space” version of the approach. Why

should changing the variable from discrete to continuous solve any problem? But if it does, where is the proof?

With all this information at hand, I can only be skeptical about the meaning of the results obtained with this approach. How can I know that the results presented in figure 2 are not just an artifact of the method? In my opinion, there are enough results from the other analyses for the authors to make their arguments on the bio-geographic processes discussed in the article. Adding the “continuous spatial diffusion” results to the article is just a risk of publishing nonsense and additional promotion of a method that has not been properly evaluated.

We thank the reviewer for their comments. We have added a passage discussing the issues that DTA and related methods face, particularly in relation to sampling bias, in the context of our study. We have added a paragraph on this issue (line 542).

In light of your comments we also contacted Philippe Lemey to ask about the issues you raised. A summary of his comments can be found below:

*The reviewer’s argumentation about DTA is correct; it is indeed more sensitive to sampling bias compared to structured coalescent approaches. And if available for / scalable to the problem under investigation, they should be preferred over DTA. However, this comparison is not relevant for the continuous diffusion model because there is no equivalent spatially-explicit coalescent implementation available (not that I am aware of). Sampling bias is likely also to be in issue, but it is less obvious than for DTA. Sampling bias may affect reconstructions (in ways that are not well characterised for the continuous model) and that like for any model, violations of its assumptions need to be considered (e.g. that Brownian diffusion implies a zero-mean displacement process).*

*We have shown that the approach can recover the true ancestral locations under a generative model that is indeed the same as used for inference. That’s about the only validation one can do. One can simulate under different generative models (i would welcome suggestions on which ones because if there are better ones, I would expect them to be implemented for inference as well), to investigate how a mismatch in assumptions may affect the reconstruction. But this is no better in terms of validation as we do not know which process will better reflect biological reality in general.*