Response to reviewers' comments

>>> We would like to thank the reviewers and the recommender for their insightful comments on our manuscript. We have incorporated almost all of their suggestions into the revised version of our manuscript, which we have uploaded onto bioRxiv (http://biorxiv.org/cgi/content/short/209254v2). Detailed responses to the comments can be found below.

Decision

by Lucy Weinert, 2017-12-13 10:37
Manuscript version: 10.1101/209254

The dynamics of preferential host switching: host phylogeny as a key predictor of parasite prevalence and distribution

This manuscript describes a new model, which takes in to account host parasite cospeciation events as well as biased host switching in the form of the "host phylogenetic distance effect" i,e. the ability of parasites to preferentially switch and establish on closely related hosts. Importantly, the model makes predictions, which should be relatively straightforward to test with comparative data.

Both reviewers thought the paper was very well written and had just minor comments to make on the structure, specifically about incorporating some of the supplementary figures in to the text. I agree with their suggestions that it is hard to make it through the results without referring to the supplement. In particular, in agreement with one of the reviewers, I very much liked the section on the importance of host trees with specific examples and wondered whether there was a way of including a cartoon schematic of Figures S2A-D without having the actual large figures moved to the main text. It would also be nice to see the final infection frequency of hosts with these trees, as to my eyes, it is not obvious.

>>> We agree that it would be nice to show those trees in the main text and have now included two of the trees in Figure 3. Since these trees are very large, we still show the same trees as well as the two other trees in the SI as full-page figures. We now mention the infection frequency in the figure caption of Fig. 3 as well as directly in the figures for Fig. S1.

One of the reviewers also has some comments on how tree imbalance metrics affect infection frequencies and some further clarification points for the discussion.

Reviews

Reviewed by Damien de Vienne, 2017-12-13 12:51

In this manuscript entitled "The dynamics of preferential host switching: host phylogeny as a key predictor of parasite prevalence and distribution", the authors propose a model to simulate the evolution of parasite species along a host

phylogeny, taking into account the phylogenetic distance effect (PDE), i.e. the fact (commonly observed) that host-shifts occur preferentially towards hosts that are closely related (phylogenetically) to their host of origin. They evaluate the impact of this PDE and other host-tree-related parameters (speciation rate, extinction rates, tree size, etc.) on the prevalence and distribution of parasites across the host species. They end up with important predictions under the model assumptions: we should see more parasite when the host trees display a few large clades than when they show many small ones; host species turnover increases parasite prevalence, and small host clades harbour less parasites than large ones. I had a great time reading this paper that I found very clear and well written. I think that it is an important piece of work that clarifies many points and will be useful for future work, for comparing observations to predictions and better understanding host-parasite dynamics in biological systems. I only have some minor points that I would like to mention. First, the authors test the effect of PDE , which is nice because it is never considered in cophylogenetic methods, but at the same time they decide to forget about classical events such as failure-to-speciate or within-host speciation. The authors consider that "they do not expect this to affect their results qualitatively" (L. 426). I wonder why they did not include that in the model from the start, given the simplicity (or am I wrong?) of considering it. This should be justified.

>>> The reason we did not include these effects was to keep the model as simple as possible. However, incorporating these effects into the simulation algorithm is indeed straightforward and we have done so now (described in detail in SI section 1.2 and 1.3). We have run additional simulations incorporating these two effects and now present the results in new figures in the SI. Since including all the results (three figures per model extension) would have inflated the SI too much, we have chosen to just show a single summary figure per model extension. In addition to results for lower and faster parasite turnover (Figs. S4 and S5), the model extensions are now shown in Fig. S6 (coinfection), S7 (parasite loss during cospeciation) and S8 (parasite speciation within a host).

Second, it was suggested and experimentally shown by de Vienne (2009) that a parasite may be better at infecting a host that its close relatives can infect, making the parasite phylogeny a predictor of host shifts as well. This should be mentioned.

>>> We now mention this interesting result in the Introduction.

Third, the author mention earlier work that estimated the impact of the shape of the tree (imbalance level) on the parasite-related parameters. Why didn't the authors look at the impact of this imbalance level (like Colless index) on the results? I think that none of the features considered relate to that, even indirectly. This may be interesting to add in order to better compare with previous work.

>>> We had originally looked at the Colless index but did not observe a correlation with infection frequencies, so we had not included these results. We have now analysed this further and found that when the Colless index is calculated not for the full host trees but for host trees where all extinct lineages have been pruned, there is a positive correlation with infection frequency in the presence of the phylogenetic

Fourth, the authors consider that host-shift and speciation are concomitant. In our 2009 paper, we showed that (with our model) congruence between host and parasite phylogenies was only obtained when the time between shift and speciation was dependent on the distance of the shift (large distance, small time to speciation). This is apparently not the case here. This could be noted.

Finally, I think that it would make sense to move figure S2 to the main text (at least one of the 4 panels), because it is discussed at length in the manuscript, and because the whole paper is on host and parasite trees and we don't see any (apart from the sketch of Figure1).

Reviewed by Nathan Medd, 2017-12-13 12:51

This manuscript introduces a stochastic model predicting the effects of host phylogenetic distance on the distribution of (specialised) parasites across host clades. The authors generate testable predictions about the distribution and frequency of infection with and without this effect. The manuscript tackles a genuine knowledge gap as highlighted by de Vienne et al. (2007). It is well written and arguments follow a logical order making its predictions and conclusions easy to interpret. The model presented is fairly simple and makes some large assumptions: the most obvious being that parasites are host specific and that speciation only occurs when host speciation does. This limits the model's predictive power for a

large swathe of parasites; however, the authors are upfront with the limitations of their model and its application to more host-specific parasites. This is not a strong criticism (I will spare you the George Box quotation).

One general comment is that there seems to be a strong reference to supplementary figures throughout the paper. I'm aware figure space is limited, but this may make reading the results a little difficult depending on the eventual format of the paper when published (some elaborated below).

>>> As mentioned above, we have now included two cophylogenies in the main text (Fig. 3C), and also added the plots for the no-PDE case that were previously in Fig. S1 to Fig. 2.

Minor Comments: - 236. Introduce abbreviation: "Most recent common ancestor (MRCA)" - 249/Figure2.

>>> Since we were using this abbreviation only once we have now omitted it entirely.

I would like to see in the main text the comparison of fig. 2 to figure S1, especially 2D vs S1D. The distribution of correlation coefficients between parasite and host with and without PDE seems like an important result that deserves to be in the main text.

>>> We have now made these plots more narrow and combined Fig. 2 and S1 into a single figure so that the PDE and no-PDE case can be compared directly. We have adjusted the main text accordingly to provide a better comparison of the two cases.

- 309/Figure 5 (y-axis). Combination of 'parasite survival' and 'number of infected hosts'. Parasites must survive to be counted as infecting at the end of the simulation but are 'probability of survival' and 'number of hosts infected' the same thing or a combined measure? Please clarify or relabel.

>>> These are two different measures: survival probability is the fraction of simulations where the parasites survived, whereas the infection frequency is the fraction of hosts that are infected at the end of the those simulations where the parasites survived. We have now changed the label on the y-axis of Fig. 5 to be explicit about this distinction.