

Plan

Shinichi Nakagawa

07/07/2018

Robert says

- Title: I like the two new ones you propose. Another possibility would be something along those lines: “Mutualists follow Fahrenholz’s Rule of congruent host-symbiont cophylogeny more closely than parasites”
- Your proposed outline for the Nature-style abstract/first paragraph seems fine to me! I’m not so sure about using Darwin’s tangled bank quote, however; it has been used and abused quite a lot already.

Is it worth breaking down the microbe result to see if this is driven by a particular microbe taxon or host-association? I’m guessing that insect-symbionts, such as defensive or trophic bacteria like *Wolbachia* are driving the relationship.

Just to check, given that vertical transmission and mutualism are probably linked, does the effect of mutualism stand over and above the effect of vertical transmission?

Regarding moving forward:

It sounds like we have a relatively clear line of approach now. If Shinichi can get me a copy of the meta-analysis methods, results, and supplementary results, I can put together a first draft to email around - does this work for you both?

I’m currently drafting the methods breakdown, I don’t think this will affect things much. One potential issue that we haven’t dealt with is that there is a different algorithm for optimising symbiont reconstructions between TreeMap 1 and TreeMap 2. TreeMap 1 uses a fairly basic algorithm to optimise cospeciations, TreeMap 2 uses the more complex ‘jungles’ algorithm which includes a slightly more advance method of modelling host switches, and applies a cost based approach to optimising the parasite tree. We could include a supplementary test to compare any differences, but I would need to go back and try to identify which version each author used - I didn’t do this originally, as a problem is that authors do not always distinguish between the version of TreeMap used, or if they applied the default or a modified cost matrix in TreeMap 2 (which is another issue that could be investigated, i.e. how does the cost matrix affect results). However, since we combine our analyses, and our focus is not on relatively minor methodological differences, I hope this shouldn’t be a particular issue, and it might be best left to the possibility that a reviewer might query it - what are your feelings on this? I’ve attached Page 1995, and Charleston 1998, which describe the approaches taken in TreeMap 1 and TreeMap 2 respectively in case you are interested.

Also, we will need to take care navigating the cospeciation vs cophylogenetic tripwire i.e. we don’t consider/can’t evaluate cospeciation, just cophylogeny - this is mainly because we don’t have date information for host and symbiont phylogenies, and so can’t test between strict cospeciation and host speciation followed by host switching, producing a pattern mimicking strict cospeciation. Basically, we just need to be careful with our language, and have careful, clear hypothesis and interpretations.

I think we can use the current draft introduction (attached) to a large extent - this was carefully thought out and edited by Robert previously. Albeit with a better title - currently we have:

“A quantitative test of the generality of host-symbiont phylogenetic congruence”

Good to start thinking of a better one. To reflect the major result, could be something along the lines of:

“A quantitative test of host-symbiont cophylogeny reveals increased congruence is linked to mutualism” or “Host-symbiont macroevolution is driven by position on the mutualism-parasitism spectrum”.

This needs more thought, and it would be great to have your suggestions.

Another key issue will be the Nature style abstract/first paragraph. There are several possible angles here. Do either of you have a favored line into the study?

One could be understanding the mechanisms that structure biological diversity, i.e.:

-Understanding the mechanisms that structure biological diversity is a major question in evolutionary ecology. -There is growing appreciation of the role of host-symbiont interactions play in life on Earth, and in the structuring of biodiversity. -Not least, due to the realisation that the major chunk of biodiversity is composed of parasites and mutualists. -Yet, amazingly we remain ignorant about what forces influence macroevolutionary patterns of these organisms. -A key question is the extent to which symbionts codiverge with their hosts. That is, to what extent are symbiont phylogenies congruent, mirroring those of their hosts, as opposed to showing relatively weak macroevolutionary associations that frequently breakdown and form anew? -Here we instigate a new era in the study of host-symbiont macroevolution, conducting a large meta-analysis consisting of the recent accumulation of studies that use host-symbiont molecular phylogenies to estimate their cophylogenetic signal. -We find that phylogenetic congruence is greater in mutualistic associations than in parasitic associations. This suggests that either bi-directional benefits, as opposed to one-sided benefits, or decreased antagonistic effects, are important for promoting macroevolutionary congruence (i.e. decreasing the incidence of host switching), between hosts and symbionts.

This idea was broadly contemplated by Darwin in the final page of *On the Origin*, and we could use it as a quote somewhere, i.e. in the general framework of helping to elucidate the laws that have formed these co-dependent organisms:

“It is interesting to contemplate a tangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent upon each other in so complex a manner, have all been produced by laws acting around us.”

Regarding a draft outline of figures for the main document and the supplementary info, please let me know what you think about this potential list of figures:

1. A photo collage seems like a good idea. I imagine the aim here would be to highlight something along the lines of the different forms of host-symbiont interaction (i.e. parasitism vs mutualism, and the various taxonomic levels among which these relationships are distributed, emphasising the general importance of this study for biological diversity), or do you have a different view for this? In the above case, I guess we need a few examples of parasitism and a few examples of mutualism (maybe three of each?). In such case, the main challenge is probably sourcing appropriate pictures for this. I don't mind about what we use, do you guys have some ideas? I can try and source an oak gall wasp picture (i.e. insect-plant parasite example), or a Strepsiptera example (insect-insect parasitism), if necessary.
2. An example of what a tanglegram is (i.e. a simple example illustrating a host-symbiont phylogeny and what we mean by phylogenetic congruence and incongruence). This could potentially be combined with figure 1 above.
3. The main result of combined methods, for mutualism vs parasitism.

Possibly also:

4. The result for symbiont type.
5. The result for transmission type. Perhaps this risks attracting more criticism during review - i.e. somewhat flakey knowledge of which taxa are vertically transmitted vs horizontally transmitted... However, potentially mitigated by just saying this is an important issue for resolution, and this is how the data look given current understanding?

N.B. the complete set of figures in Supplementary Information.

More from Alex

Our methods (TreeMap and Parafit), test the null hypothesis that parasite phylogeny is independent of host phylogeny, by generating random phylogenies (with the same number of taxa and the same hosts), and measuring fit. The proportion of random trees that have the same (or greater) fit, equals the probability of obtaining the observed value due to chance alone.

For each study we know what the lowest probability achievable is, given the constraint imposed by limiting the number of randomizations performed, i.e. for 100 randomizations it is 1/100 or 0.01, for 1,000 it is 0.001, for 10,000 it is 0.0001. In cases of poor observed phylogenetic congruence, the result does not reach the limit, achieving a probability above it, so the limit is not a major issue. Where authors see high or perfect phylogenetic congruence, the limit becomes an issue, as the true probability is typically somewhere between the limit and zero. Thus, the randomisation limit leads to an underestimation of the true effect size. The question is, how bad is an underestimation of the true effect size statistically speaking?

Presumably similar problems exist in for example, quantifying disease infection, where a negative result among a certain sample size (say 1,000 individuals), cannot be taken as the absence of infection. The larger the sample size, the more accurate the quantification of prevalence, but the exact figure cannot be known unless the entire population is sampled. In our case, the total population of possible rooted phylogenetic trees for n symbiont taxa is: $(2n-3)!/(2^{n-2}(n-2)!)$. But, unfortunately we do not know how many among this set of potential symbiont trees show similar or greater phylogenetic congruence with the host tree, as the particular symbiont tree we are testing.

I'm now back in the UK, but from the 13th July I'll be collecting butterflies in Transylvania for two weeks, and then will be in Sweden again for two weeks. I should be available by email and should have some time to work while I'm away. But after this I'm done with travel for the foreseeable future, and getting this project submitted is a key priority. Given the novelty and interesting findings I remain hopeful that we can get the manuscript accepted in a high ranking journal. Happy to Skype or discuss the issue more as useful.

Shinichi's thoughts