False detection of "true" species under the multi-species coalescent model

Posted on 7 February, 2017 by Katie Everson

The multi-species coalescent model (MSCM) is the biggest name in the game (if the game is genetic species delimitation). But a new paper from *Proceedings of the National Academy of Sciences* asks: is the MSCM really doing what we think it's doing?

Some background: The MSCM, usually implemented in the program BPP (Yang & Rannala 2010), models speciation as an instantaneous event under the birth-death process.

But we know that the biological reality is more complex. Within most species there is some amount of gene flow restriction (e.g., due to environmental or geographic barriers), not all of which will eventually lead to speciation. Depending on the extent and duration of isolation and the strength of selection, speciation can be a gradual and stochastic process.

Sukumaran & Knowles (2017) tested the performance of the MSCM using data simulated under the "protracted speciation model," which includes a few more biologically relevant parameters compared to the simpler birth-death model. Two key components of the protracted speciation model are the species conversion rate (c, the rate at which incipient species develop into true species), and another parameter that accounts for incipient species going extinct or merging back into their parent species.

The authors used two different simulation schemes: a "fixed duration" scheme, where the simulations ran for a fixed amount of time and produced varying numbers of species, and a "fixed species number" scheme, where simulations ran until five species were generated.

Perhaps you've already guessed what happened: Sukumaran & Knowles found that the MSCM is great at identifying lineages, but it overestimates the number of species. In fact, the MSCM can estimate 5 to 13 times more than the true number of species. It is also worth noting that the errors are all positive; i.e., BPP never underestimated the number of true species but only overestimated them.

Why does it matter? These methods lead to inflated diversity estimates, with direct consequences for conservation and ecology research. For now, the authors suggest using morphological, ecological, ethological, or other classes of data to correctly attribute MSCM results to either species-level or population-level processes – a call that has been echoed by other researchers in the last 6 months (e.g., Freudenstein et al. 2016).
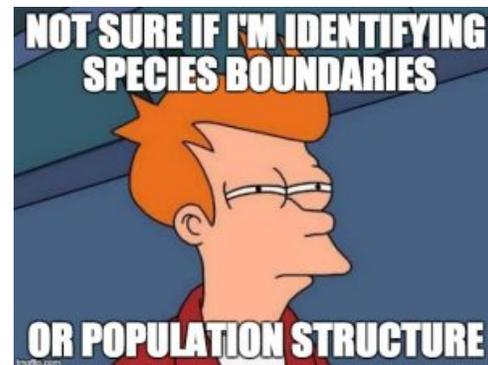
This study also served as a call for new methods for genetic species delimitation, and the researchers have already tweet-hinted at a new method that may be coming down the pipe soon. I imagine the new method will have some basis in protracted speciation model? I'm looking forward to reading it.

Cited:

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, 201607921.
doi: 10.1073/pnas.1607921114

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, *107*(20), 9264-9269. doi: 10.1073/pnas.0913022107

Freudenstein, J. V., Broe, M. B., Folk, R. A., & Sinn, B. T. (2016). Biodiversity and the Species Concept—Lineages are not Enough. *Systematic Biology*. doi: 10.1093/sysbio/syw098
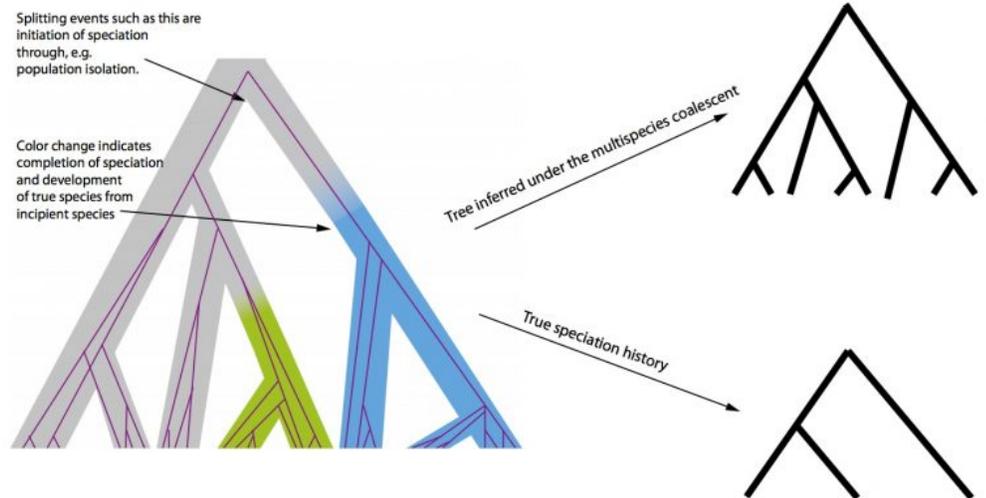

Figure 1 from Sukumaran & Knowles 2017. The multi-species coalescent model may over-estimate the "true" number of species on a phylogeny.
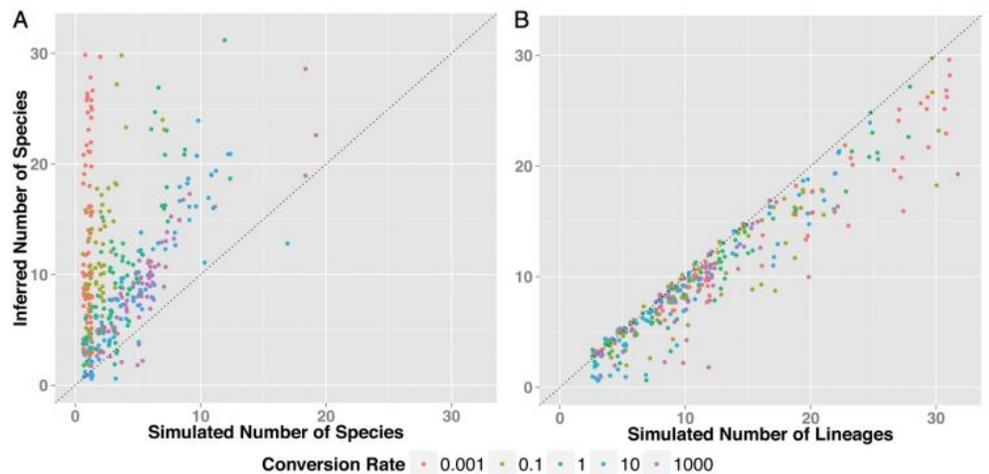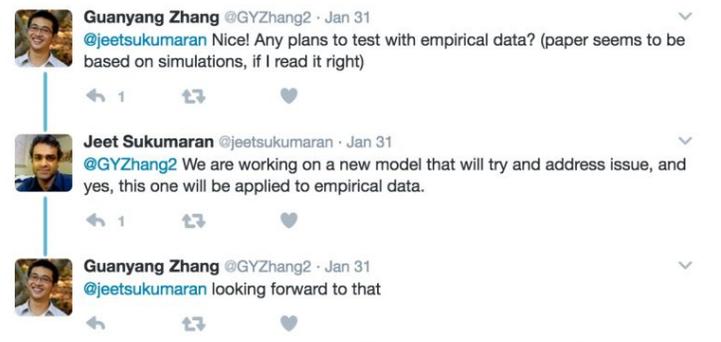

Figure 2 from Sukumaran & Knowles 2017. The multi-species coalescent model infers more than the true number of species, under a variety of simulation conditions (A). However, it does correctly estimate the number of lineages (B).

About Katie Everson

I'm a PhD candidate at the University of Alaska Museum studying the evolution of Madagascar's tenrecs. Alaska is really far from Madagascar -- that's why I love museum collections! My core research interests are phylogeography and species delimitation.

View all posts by Katie Everson →

This entry was posted in software, speciation and tagged BPP, modeling, multi-species coalescent, software, speciation, species delimitation. Bookmark the permalink.

0 Comments   **The Molecular Ecologist**   1   Login

♥ Recommend   ⤴ Share   Sort by Oldest

Start the discussion…

Be the first to comment.

✉ Subscribe   Ⓓ Add Disqus to your site Add Disqus Add   🔒 Privacy

Proudly powered by WordPress.