Statistical Comparison of Trait-Dependent Biogeographical Models Indicates That Podocarpaceae Dispersal Is Influenced by Both Seed Cone Traits and Geographical Distance

KRISTINA V. KLAUS[1,*] AND NICHOLAS J. MATZKE[2,3]

We note that it has recently been claimed that "DEC and DEC+J cannot be compared using standard statistical methods" (Ree and Sanmartín 2018). However, this argument can be disproved on several fundamental grounds. For example, the identifiability of the d and j parameters was proven with simulation-inference experiments in Matzke (2014), whereas Ree and Sanmartín omitted simulation and instead argue from 2 small, human-constructed datasets (trees with 2 or 4 tips) – which we consider an insufficient basis on which to judge Maximum Likelihood inference of models with 2 or 3 free parameters. We briefly review these issues in Supplementary Material available on Dryad. A more detailed response by Matzke is in preparation; see also McDonald-Spicer et al. (2019).

Appendix B. Preliminary response to the Ree/Sanmartín critique of statistical comparison of the DEC/DEC+J models

We note that it has recently been claimed that "DEC and DEC+J cannot be compared using standard statistical methods" (Ree and Sanmartín 2018). However, this argument can be rejected on several preliminary grounds that we review briefly; a more detailed response by Matzke is in preparation. First, Ree and Sanmartín totally ignore the extensive simulations presented in Matzke (2014) demonstrating the distinguishability of the two models and the identifiability of the d and j parameters under a variety of conditions. Second, the claimed problem (the jump-dispersal relative weight parameter j approaching its maximum allowed value) is almost never observed in empirical datasets (the exception is the rare case where all species in the phylogeny are allopatric, a situation in which an exclusive role for an allopatry-promoting process like jump dispersal is reasonable; McDonald-Spicer et al. 2019). Third, the authors' only examples of unintuitive inference are two tiny, hand-constructed datasets (phylogenies of 2- and 4-taxa), datasets too small to expect reliable inference for a Maximum Likelihood method operating on models with 2 or 3 free parameters; these datasets are too small to even calculate AICc, for instance, as would be obligatory for any statistical model comparison on such small datasets. The valid test of an inference method is simulation-inference experiments, not individual human-constructed datasets. Finally, it is extremely hazardous to allow large differences in likelihood between directly comparable models (DEC and DEC+J differ by just one free parameter, and DEC is a special case of DEC+J) to be casually disregarded: differences in likelihood are fundamental not just to AIC-based approaches but also to frequentist (e.g. the Likelihood Ratio Test) and Bayesian statistical inference generally, and underlie vast amounts of research in evolutionary biology and dozens of other fields.

Further relevant points, for example, validating the DEC and DEC+J likelihoods by calculating over all possible datasets, and showing that comparing DEC and DEC+J is identical to comparing two special cases of ClaSSE (with parameters set to match a pure-birth, Yule process for the tree) have been available from June 2017 and June 2015, respectively, at the online BioGeoBEARS help website (PhyloWiki), specifically the BioGeoBEARS Validation page at: http://phylo.wikidot.com/biogeobears-validation.