

Comments on Mayrose et al.

Graham Jones

2015-05-07

SYSTEMATIC BIAS IN BiSSE

To test the approach of Mayrose et al. (2011), we simulated trees with similar characteristics to the empirical trees and then followed their method for assessing differences in divergence rates. The analysis was carried out in R (R Core Team, 2013) using the diversitree package which implements the BiSSE model (Maddison, 2007). All of the R scripts used for this paper are in the supplementary material. In the simulations, both the diploid speciation rate λ_D and the polyploid speciation rate λ_P were set to 0.2, and the corresponding extinction rates μ_D and μ_P were both set to 0.1. The transition rate q_{DP} was set to 0.02. The trees were allowed to grow for a maximum of 26 time units. Trees with fewer than ten tips, and those with diploids only or polyploids only at the tips, were rejected. These settings were found to produce trees with a similar range of sizes to those in the empirical data set of Mayrose et al. (2011), with about one third of the tips being polyploid. The simulated trees were then analyzed using the methods of Mayrose et al. (2011), using a Bayesian approach and the Markov chain Monte Carlo (MCMC) method described in FitzJohn et al. (2009).

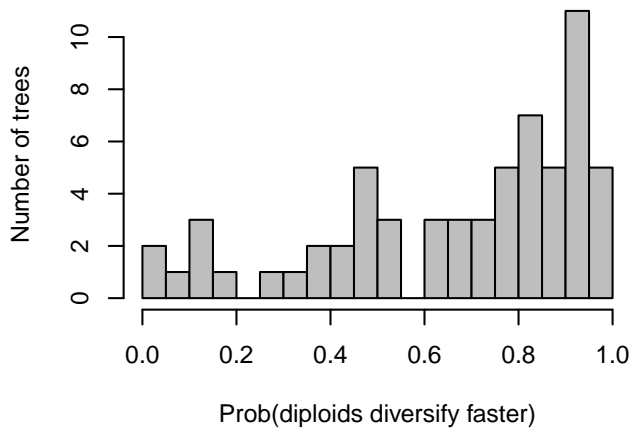


Figure 1: Posterior probabilities that diploids exhibit higher diversification rates, obtained from simulated trees in which no bias exists.

The main result is shown in Figure 1, which is directly comparable to Figure 1A of Mayrose et al. (2011). In our case, however, the bias towards one (high posterior probabilities for diploids having higher diversification rates) is a pure artifact. The reason for the bias is not yet fully understood. However, the situation is clearly asymmetric between diploids and polyploids: diploids can become polyploids, but not vice versa, and there are more diploids than polyploids in the set of trees. The supplementary material contains two examples of how bias can occur in similar, but simpler situations.

Mayrose et al. (2014) argue that “approaches like... [theirs] ...that consider the preponderance of evidence across multiple clades is the best way to assess whether a trait, like polyploidy, affects diversification in a repeatable way.” A meta-analysis of the type conducted by Mayrose et al. (2011) is indeed attractive. It can reduce the variance of individual analyses so that a useful signal

can be revealed. However, it is vulnerable to a systematic bias affecting all the individual analyses, as our results show. Used on a single tree, BiSSE is unlikely to produce misleading results. The bias will usually be overwhelmed by uncertainty. Figures S1 and S2 show that the bias diminishes with tree size, and other results with larger trees (such as Figure S3, supplementary material) confirm this. The results of meta-analyses using BiSSE are likely to be affected by bias if the true transition rates are significantly different and the tree sizes are fairly small, as in the case of the empirical data analyzed by Mayrose et al. (2011).

In support of their argument, Mayrose et al. (2014) present data from Goldberg and Iqic (2008), indicating in Figure 1 that speciation rate estimates are centered near the true values, and report, but do not show, similar results for extinction rates. These results appear at first sight to be in conflict with our results, so some explanation seems warranted.

In Figure 1 of Mayrose et al. (2014), there is no information about the number of terminals (species) in the trees of the Goldberg and Iqic (2008) simulations. However, in all but the center and left panels of the bottom row of the figure, which have 55 and 56 terminals, respectively, the number of terminals is over 100. The genera included in Mayrose et al. (2011) have a mean of only 38.7 terminals, and over half have fewer than 30 terminals. The estimates in Figure 1 of Mayrose et al (2014) show a large uncertainty in these two panels, and it is not clear whether this hides a systematic bias. Furthermore, Mayrose et al. (2014) do not specify what measure of centrality they use. Because some of the distributions are quite skewed, a median and mean (for example) could give different results.

In the supplementary material we show results that shed more light on the issue. In Figure S1 we show posterior means of the parameters $\lambda_D, \lambda_P, \mu_D, \mu_P, q_{DP}$, and of the difference $r_D - r_P$ between the diploid divergence rate $r_D = \lambda_D - \mu_D$ and the polyploid divergence rate $r_P = \lambda_P - \mu_P$, plotted against tree size. The estimates for λ_D, λ_P , and μ_D are not noticeably biased, but μ_P, q_{DP} , and $r_D - r_P$ are clearly overestimated. In Figure S2 we show maximum likelihood estimates (MLEs) of the same quantities. These estimates have much wider spread than the posterior means. The results for speciation rates and for μ_D are similar in general form to those in Figure 1 of Mayrose et al. (2014), but others are not. The distributions of the MLEs of μ_P and q_{DP} are extremely diffuse and highly skewed. For example, estimates of μ_P can exceed 2.0, more than twenty times its true value of 0.1, while around a quarter of the estimates are less than 0.0001. The median of the MLEs of μ_P is 0.094, a little low, but the mean is 0.211, over double the true value. The only safe conclusion we can draw is that MLEs of μ_P and q_{DP} are almost useless for trees of this size.

References

- R. G. FitzJohn, W. P. Maddison, S. P. Otto (2009) Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* 58, 595 (2009). doi:10.1093/sysbio/syp067 Medline
- Maddison, Wayne P. and Midford, Peter E. and Otto, Sarah P. (2007) Estimating a Binary Character's Effect on Speciation and Extinction. *Systematic Biology*, **56**:(5), 701–710.
- Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. (2011) Recently formed polyploid plants diversify at lower rates. *Science*, **333**, 1257.
- Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2014) Soltis, D. E., M. C. S. Salcedo, I. J. Thaden, L. M. Majure, N. M. Miles, E. V. Mavrodiev, W. Mei, M. B. Cortez, P. S. Soltis, and M. A. Gitzendanner. *Phytologist*.
- R: A language and environment for statistical computing. R Foundation for Statistical Computing, R Core Team (2013) Vienna, Austria. URL <http://www.R-project.org/>.