

Parsimony and maximum-likelihood phylogenetic analyses of morphology do not generally integrate uncertainty in inferring evolutionary history: a response to Brown *et al.*

Mark N. Puttick^{1,2}, Joseph E. O'Reilly¹, Derek Oakley¹, Al R. Tanner¹, James F. Fleming¹, James Clark¹, Lucy Holloway¹, Jesus Lozano-Fernandez¹, Luke A. Parry¹, James E. Tarver¹, Davide Pisani^{1,3}, Philip C. J. Donoghue¹

¹School of Earth Sciences and ³School of Biological Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK

²Department of Life Sciences, The Natural History Museum, Cromwell Road, South Kensington London SW7 5BD, UK

Our recently study evaluated the performance of parsimony and likelihood-based models of phylogenetic inference based on categorical data [1]. We found that a Bayesian implementation of a likelihood model produced more accurate results than either of the competing implementations of parsimony (the main method currently employed) and the Maximum Likelihood implementation of the same likelihood model, principally because the results of Bayesian analyses are less resolved (less precise) because this MCMC-based approach intrinsically integrates uncertainty in topology estimation. Of the three main methods, Maximum Likelihood performed the worst of all.

In their comment on our article, Brown and colleagues [2] argue that our experiments are invalid because we did not employ methods of measuring uncertainty associated with the Maximum Likelihood method. When bootstrapping is employed, the 50% consensus tree derived from Maximum Likelihood is comparable to the result of the Bayesian analyses, as summarised by its majority rule consensus tree. On this basis, Brown and colleagues conclude that they cannot advocate one method of phylogenetic inference over another: Bayesian, Maximum Likelihood and parsimony methods differ and thoughtful consideration is required in order to choose among these methods. Unfortunately, their analyses do not support this conclusion since their focus is only on the performance of the two implementations of the same likelihood model, not on their performance relative to parsimony, which was the focus of our study - that, following our and others previous studies [1, 3, 4], roundly rejects parsimony in favour of the likelihood inferential framework.

The principle thrust of the argument presented by Brown and colleagues [2] is that the experiments performed by Puttick et al. [1] did not allow for a fair comparison between phylogenetic methods: the Bayesian implementation intrinsically integrates uncertainty, while it is common practise to evaluate uncertainty *post hoc* for Maximum Likelihood and parsimony inference using bootstrap methodology. In Puttick et al. [1], we explicitly addressed this issue in two ways. The first argument is that bootstrapping is not an intrinsic aspect of Maximum Likelihood estimation or parsimony phylogenetic analysis. Thus, we did not need to consider support values in our analyses. Using Bayesian estimation, it is intractable to analytically estimate topology using the Mk model and so it is necessary to use a MCMC procedure to produce an un-normalised posterior distribution of trees. From this posterior distribution, it is straightforward to interpret a 50% majority-rule consensus tree and clade support measures (posterior probabilities), unlike analogous measures produced from bootstrapping [5]. Our second argument was that bootstrapping is unsuited to analysis of categorical data because its expectations are not met, viz. that the phylogenetic signal is not independently and identically distributed through the data, which is a view common to phylogenetic textbooks, for example Felsenstein [6], Kitching et al. [7], and Schuh [8]. Brown et al. highlight that this is an issue shared by both Bayesian and Maximum likelihood implementations of the Mk model, as independence is assumed when calculating the likelihood of individual characters. When bootstrapping, the assumption of independence is violated more times than when the Bayesian

implementation of the Mk model is used (once in the original Maximum likelihood estimation, and then once further in the non-parametric bootstrapping procedure). Furthermore, the interpretation of posterior probabilities as the probability of observing a clade given the morphological data is straightforward, whereas the exact meaning of a bootstrap proportion is still equivocal, with numerous proposed interpretations [9].

We concede that, despite our concerns, bootstrapping has been used commonly in phylogenetic reconstruction, including phylogenetic analyses based on morphological traits. In this sense, our experiments could be viewed as failing to faithfully simulate common practise. However, while it is common practise to measure support for topologies through bootstrapping and Bremer Support in Maximum Likelihood and parsimony phylogenetic analyses of morphological traits, most studies present support measures on more fully resolved trees that include nodes with negligible support, rather than collapsing nodes that exhibit less than 50% support, as Brown and colleagues suggest [2]. To underline the prevalence of this approach we reviewed studies citing Lewis [10], the originator of the Mk model, published since the start of this year, as recorded in Web of Science (census date XX June 2017). Of the 48 citing articles (See ESM/ Table), 31 phylogenetic studies were based on morphological traits, in whole or in part. Of the 11 studies that employed Maximum Likelihood, 10 evaluated bootstrap support, all of which resolved nodes with <50% support. The same approach is seen in parsimony analyses where, among 18 studies, only 12 evaluated bootstrap support, of which, 8 resolved nodes with <50% support - though these nodes were usually supported by other metrics like Bremer Support. Resolution of unsupported nodes is less prevalent in Bayesian analyses where, among the 29 studies examined (27 of which presented posterior probabilities), only 12 resolved unsupported nodes; many of these were in Maximum Clade Credibility trees. Unsupported nodes were present in Bayesian trees in only 2 of the 9 studies that employed both Maximum Likelihood and Bayesian analysis. Thus, while many of these studies present Maximum Likelihood and parsimony-based trees that are more fully resolved than their support measures should perhaps permit, when they are associated with parallel Bayesian analyses, these are invariably summarised by majority rule consensus.

Hence, the experiments presented in Puttick et al. [1] followed common practise, as exemplified by the literature. Brown and colleagues [2] are correct in their view that measures of support are widely employed in phylogenetics. However, most Maximum Likelihood- and parsimony-based studies effectively ignore *post hoc* topological support measures in their inferences of evolutionary history, which are most often based on more fully resolved, Maximum Likelihood and parsimony, trees. Practise shows that the same is not true of Bayesian analyses which are invariably summarised by the intrinsic majority rule consensus (though some studies also seek further resolution using other methods for summarising a distribution of trees, such as Maximum Clade Credibility). While we agree with Brown and colleagues [2], that poorly supported clades should be collapsed in Maximum Likelihood or maximum parsimony topologies, this does not reflect the ways in which workers actually use Maximum Likelihood and, indeed, parsimony, in the phylogenetic analysis of morphological trait data.

In effect, Brown and colleagues [2] have not addressed the core questions of our study. Rather, they have extended the experiments we undertook, with a different aim, and they have extended the conclusions. They observe that when clade support is considered, Maximum Likelihood and Bayesian implementations of the Mk model perform equally well. This is an important observation that will provide some confidence in Maximum Likelihood-based analyses of morphological trait data - just as soon as common practise catches up with the need to control for topological uncertainty in inferring evolutionary history.

Brown and colleagues [2] close out their manuscript without advocating a method of phylogenetic inference and, indeed, argue that there is no superior method. Suitable methods, they argue, should be identified in each instance given the biological question at hand. In so doing, they explicitly draw parsimony back into consideration – despite the fact that their analyses do not address this method. This populist view ignores previous studies that highlight the inaccuracy of parsimony [1, 23, 24], to which they present no counter-evidence. This was the focus of our study. Our experimental design, focussed on common practise, is valid, as are the results, interpretations and conclusions that we derived from our experiments.

References

1. Puttick M.N., O'Reilly J.E., Tanner A.R., Fleming J.F., Clark J., Holloway L., Lozano-Fernandez J., Parry L.A., Tarver J.E., Pisani D., et al. 2017 Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society B: Biological Sciences* **284**(1846). (doi:10.1098/rspb.2016.2290).
2. Brown J.W., Parins-Fukuchi C., Stull G.W., Vargas O.M., Smith S.A. 2017 Bayesian and likelihood phylogenetic reconstructions of morphological traits are not discordant when taking uncertainty into consideration. A comment on Puttick et al. *Proceedings of the Royal Society B: Biological Sciences*. (doi:<https://doi.org/10.1101/114793>).
3. O'Reilly J.E., Puttick M.N., Parry L.A., Tanner A.R., Tarver J.E., Fleming J., Pisani D., Donoghue P.C.J. 2016 Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters* **12**, 20160081. (doi:10.1098/rsbl.2016.0081).
4. Wright A.M., Hillis D.M. 2014 Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS One* **9**(10), e109210. (doi:10.1371/journal.pone.0109210).
5. Yang Z., Rannala B. 2005 Branch-length prior influences Bayesian posterior probability of phylogeny. *Systematic Biology* **54**(3), 455-470. (doi:10.1080/10635150590945313).
6. Felsenstein J. 2004 *Inferring phylogenies*. Sunderland, MA, Sinauer; 664 p.
7. Kitching I.J., Forey P.L., Humphries C.J., Williams D.M. 1998 *Cladistics: the theory and practice of parsimony analysis*. 2nd ed. Oxford, Oxford University Press; 228 p.
8. Schuh R.T. 2000 *Biological systematics: principles and applications*. Ithaca, NY, Constock; 236 p.
9. Yang Z. 2006 *Computational molecular evolution*. Oxford, Oxford University Press.
10. Lewis P.O. 2001 A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**(6), 913-925.