PROCEEDINGS B

rspb.royalsocietypublishing.org

Invited reply

Cite this article: Puttick MN *et al*. 2017 Parsimony and maximum-likelihood phylogenetic analyses of morphology do not generally integrate uncertainty in inferring evolutionary history: a response to Brown *et al*. *Proc. R. Soc. B* **284**: 20171636.
http://dx.doi.org/10.1098/rspb.2017.1636

Received: 21 July 2017
Accepted: 11 September 2017

Subject Category:
Palaeobiology

Subject Areas:
evolution, palaeontology, taxonomy and systematics

Authors for correspondence:
Davide Pisani
e-mail: davide.pisani@bristol.ac.uk
Philip C. J. Donoghue
e-mail: phil.donoghue@bristol.ac.uk

The accompanying comment can be viewed at http://dx.doi.org/10.1098/rspb.2017.0986.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3888106.

THE ROYAL SOCIETY PUBLISHING

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[1,3], Joseph E. O'Reilly[1], Derek Oakley[1], Alistair R. Tanner[2], James F. Fleming[1], James Clark[1], Lucy Holloway[1], Jesus Lozano-Fernandez[1], Luke A. Parry[1], James E. Tarver[1], Davide Pisani[1,2] and Philip C. J. Donoghue[1]

[1]School of Earth Sciences, and [2]School of Biological Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK
[3]Department of Life Sciences, The Natural History Museum, Cromwell Road, South Kensington, London SW7 5BD, UK

MNP, 0000-0002-1011-3442; JEO, 0000-0001-9775-253X; JC, 0000-0003-2896-1631; LH, 0000-0003-1603-2296; JL-F, 0000-0003-3597-1221; LAP, 0000-0002-3910-0346; PCJD, 0000-0003-3116-7463

Our recent study evaluated the performance of parsimony and probabilistic models of phylogenetic inference based on categorical data [1]. We found that a Bayesian implementation of a probabilistic Markov model produced more accurate results than either of the competing parsimony approaches (the main method currently employed), and the maximum-likelihood implementation of the same model. This occurs principally because the results of Bayesian analyses are less resolved (less precise) as a measure of topological uncertainty is intrinsically recovered in this MCMC-based approach and can be used to construct a majority-rule consensus tree that reflects this. Of the three main methods, maximum likelihood performed the worst of all as a single exclusively bifurcating tree is estimated in this framework which does not integrate an intrinsic measure of support.

In their comment on our article, Brown *et al*. [2] argue that our experiments are invalid because we did not employ uncertainty measures after obtaining a maximum-likelihood estimate of the topology. When bootstrapping is employed, a 50% consensus tree constructed from the bootstrap distribution is often indistinguishable from the majority-rule consensus tree constructed from the posterior sample obtained in a Bayesian analysis. This result is not entirely unexpected, as the maximum-likelihood and Bayesian statistical frameworks share many statistical similarities, including a dependence on a likelihood function that incorporates the Mk model in this context. 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 wholly support this conclusion because they exclusively focus on the performance of the two implementations of the same probabilistic model, without considering their performance relative to parsimony. This was a key aspect of our study comparing the primary methods of phylogenetic reconstruction as they are commonly implemented. Our and other previous studies [1,3,4] reject parsimony in favour of a Bayesian MCMC framework in which uncertainty is incorporated, further drawing into question the veracity of Brown and colleagues' assertion that there is equivalent performance among methods.

The principle thrust of the argument presented by Brown *et al*. [2] is that the experiments performed by us [1] did not allow for a fair comparison between phylogenetic methods: the Bayesian implementation intrinsically

integrates uncertainty, while it is common practice to evaluate uncertainty *post hoc* for maximum-likelihood and parsimony inferences using bootstrap methodology. In our study [1], we explicitly addressed this issue in two ways. The first argument was 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 an MCMC sampling procedure to produce a posterior sample of trees. From this approximation of the 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 arguably unsuited to analysis of morphological data because its statistical 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 (e.g. [6–8]). Brown *et al*. [2] correctly highlight that this is an issue shared by both Bayesian and maximum-likelihood implementations of the Mk model, as independence is assumed when summing the log-likelihood of individual characters. However, 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], all of which are contingent on the maximum-likelihood estimate of topology.

We agree that bootstrapping has been used commonly in phylogenetic reconstruction, including analyses based on morphological traits, to assign a level of support to the constituent nodes of a most parsimonious or maximum-likelihood topology estimate. In this sense, our experiments could be viewed as failing to faithfully simulate common practice. However, while it is common practice to measure support for the clades through bootstrapping in maximum-likelihood and parsimony phylogenetic analyses of morphological traits, most studies present these support measures on fully resolved topology estimates that include nodes with negligible support, rather than collapsing nodes that exhibit less than 50% support into soft polytomies, as Brown *et al*. 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 14 June 2017). Of the 48 citing articles (see the electronic supplementary material), 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 less than 50% support. The same pattern is seen in parsimony analyses where, among 18 studies, only 12 evaluated bootstrap support, of which eight resolved nodes with less than 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 two of the nine studies that employed both maximum-likelihood and Bayesian analyses. 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 summarized by majority rule consensus.

Hence, the experiments presented in our paper [1] followed common practice, as demonstrated by the literature. Brown *et al*. [2] are correct in their view that measures of support are widely employed in phylogenetics, and poorly supported clades should be collapsed in maximum-likelihood or maximum parsimony topologies. 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. Practice shows that the same is not true of Bayesian analyses, which are usually summarized by the majority rule consensus (though some studies also seek further resolution using other methods for summarizing a distribution of trees, such as maximum clade credibility). Therefore, based on current use of phylogenetic models, our support for Bayesian inference is validated based on the current practice used by phylogeneticists.

In effect, Brown *et al*. [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 practice catches up with the need to control for topological uncertainty when inferring evolutionary history.

Brown *et al*. [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 declaration ignores previous studies that highlight the inaccuracy of parsimony [1,3,4], to which they present no counter-evidence. The focus of our study was an objective comparison of the efficacy of the primary methods of phylogenetic reconstruction, including parsimony, as commonly implemented by practitioners. Our experimental design, focused on such common practices, is valid, as are the results, interpretations and conclusions that we derived from our experiments.

References

1. Puttick MN *et al.* 2017 Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proc. R. Soc. B* **284**, 20162290. (doi:10.1098/rspb.2016.2290)
2. Brown JW, Parins-Fukuchi C, Stull GW, Vargas OM, Smith SA. 2017 Bayesian and likelihood phylogenetic

reconstructions of morphological traits are not discordant when taking uncertainty into consideration: a comment on Puttick *et al*. *Proc. R. Soc. B* **284**, 20170986. (doi:10.1098/rspb.2017.0986)

3. O'Reilly JE, Puttick MN, Parry LA, Tanner AR, Tarver JE, Fleming J, Pisani D, Donoghue PCJ. 2016 Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biol. Lett.* **12**, 20160081. (doi:10.1098/rsbl.2016.0081)
4. Wright AM, Hillis DM. 2014 Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS ONE* **9**, e109210. (doi:10.1371/journal.pone.0109210)
5. Yang Z, Rannala B. 2005 Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst. Biol.* **54**, 455–470. (doi:10.1080/10635150590945313)
6. Felsenstein J. 2004 *Inferring phylogenies*. Sunderland, MA: Sinauer.
7. Kitching IJ, Forey PL, Humphries CJ, Williams DM. 1998 *Cladistics: the theory and practice of parsimony analysis*, 2nd edn. Oxford, UK: Oxford University Press.
8. Schuh RT. 2000 *Biological systematics: principles and applications*. Ithaca, NY: Constock.
9. Yang Z. 2006 *Computational molecular evolution*. Oxford, UK: Ocford University Press.
10. Lewis PO. 2001 A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925.