



Divergent maximum-likelihood-branch-support values for polytomies



Mark P. Simmons^{a,*}, Andrew P. Norton^b

^a Department of Biology, Colorado State University, Fort Collins, CO 80523-1878, USA

^b Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80523, USA

ARTICLE INFO

Article history:

Received 8 October 2013

Revised 9 January 2014

Accepted 21 January 2014

Available online 4 February 2014

Keywords:

Bootstrap

False positive

GTRCAT model

Hard polytomy

Random seed

SH-like aLRT

ABSTRACT

We applied simple 4-taxon simulations with 3-way character conflict or a hard polytomy to check for false positive branch support, with a focus on the bootstrap and recently introduced likelihood-based phylogenetic-inference programs. Given that there are only three possible bifurcating topologies, discrepancies among methods identified in this study should generally be restricted to factors other than topological search heuristics. Our four major conclusions are as follows. First, Bayesian MCMCMC posterior probabilities are not the only means of quantifying support that can produce dramatically inflated values when applied to cases of strong character conflict. Rapid bootstrapping with the GTRCAT model in RAXML can provide still greater support values for polytomies and we suggest that it generally be avoided. Second, the SH-like approximate likelihood-ratio test outperforms the bootstrap when applied to polytomies. We suggest that the SH-like aLRT be widely applied to likelihood-based empirical studies to complement the bootstrap by collapsing those branches with an SH-like aLRT percentage of ≤ 10 , irrespective of how high the likelihood bootstrap support is. Third, the 70% bootstrap cutoff does not equate to a 5% error rate and we suggest that the idea that $\geq 70\%$ bootstrap generally equates to 95% probability of accuracy in empirical analyses finally be abandoned. Fourth, rapid bootstrapping with the GTRCAT model in RAXML can generate values with very low precision, which reinforces our assertion that this method should be avoided, let alone be entirely relied upon for phylogenetic inference.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Some means of quantifying branch support is incorporated into almost every published phylogenetic analysis. The most common way of quantifying branch support for frequentist phylogenetic analyses is to use the bootstrap (Felsenstein, 1985). The bootstrap can be implemented in multiple different ways. Hence bootstrap values generated by one program and one set of tree-search settings can be dramatically different from those generated by a different program and/or different set of search settings – even when the same optimality criterion and model (when applicable) is applied. Artificially inflated bootstrap values may be caused by use of the frequency-within-replicates rather than the strict-consensus bootstrap (Davis et al., 1998; De Laet et al., 2004; Goloboff and Pol, 2005), saving only a single optimal tree per pseudoreplicate when there are multiple equally optimal trees (Goloboff and Farris, 2001; Simmons and Freudenstein, 2011; Simmons and Goloboff, 2013), and/or extrapolation of branch lengths among character partitions in the presence of non-randomly distributed

missing or inapplicable data (Lemmon et al., 2009; Simmons, 2012a, 2012b). Alternatively, artificially deflated bootstrap values may be caused by performing low quality tree searches that do not find the optimal trees in each pseudoreplicate (Freudenstein et al., 2004; Freudenstein and Davis, 2010).

Irrespective of how the bootstrap is implemented and how inflated or deflated those support values may be, numerous pseudoreplicates are required to obtain precise bootstrap values. Between 100 and 1,000 bootstrap pseudoreplicates are typically implemented in contemporary empirical phylogenetic analyses (e.g., Bacon et al., 2013; Ceccarelli and Zaldivar-Riveron, 2013; Guo et al., 2013; Kwek et al., 2013). Hedges (1992) noted that, based on the binomial distribution (also see Efron et al. (1996)), still more bootstrap replicates (1825) are required to have a 95% confidence interval within 1% for bootstrap supports of 95%. Furthermore, Hedges (1992) noted that the breadth of the confidence interval for any given number of bootstrap pseudoreplicates varies depending on the bootstrap support, with lower bootstrap supports having wider confidence intervals.

Pattengale et al. (2009, 2010) criticized Hedges' (1992) and Efron et al.'s (1996) approach to estimating the “accuracy” (properly precision; Hillis and Bull, 1993) of bootstrap values based strictly on the binomial distribution. They asserted that other

* Corresponding author. Address: Department of Biology, 200 West Lake Street, Colorado State University, Fort Collins, CO 80523-1878, USA. Fax: +1 970 491 0649.

E-mail address: psimmons@lamar.colostate.edu (M.P. Simmons).

factors (e.g., alignment quality, percentage of gaps, and strength of phylogenetic signal) also affect the precision of bootstrap values and hence the number of bootstrap pseudoreplicates performed on a given matrix should be determined using an adaptive stopping criterion for each individual matrix rather than an a priori defined number as advocated by [Hedges \(1992\)](#). [Pattengale et al. \(2010\)](#) asserted that [Hedges' \(1992\)](#) estimate of the precision of bootstrap values be used as an upper bound and that many fewer pseudoreplicates are necessary in practice.

In addition to considering whether or not the bootstrap values are inflated and precisely estimated, one needs to determine how to interpret the values. [Hillis and Bull \(1993, p. 191, 187\)](#) reported that, “bootstrap proportions [$>50\%$] provide conservative estimates of accuracy under many conditions” and “that estimated internal branches with bootstrap proportions above 70% represent true clades over 95% of the time (for the conditions tested in these simulations).” [Hillis and Bull \(1993\)](#) appropriately qualified the generality of their results and noted that their simulation conditions that produced this result are favorable to accurate phylogenetic inference.

[Hillis and Bull \(1993\)](#) is the basis for the widespread approach of using 70% as a meaningful threshold for bootstrap values. Based on a search conducted in Web of Science[®] on 30 September 2014, [Hillis and Bull \(1993\)](#) has been cited 2123 times (the second most highly cited paper from *Systematic Biology*). The large majority of these citations are for using $\geq 70\%$ bootstrap as an indication of strong support (or “well supported,” “good support,” “significant support,” etc.). For example, of the 41 empirically focused (as opposed to conceptually focused) papers published in *Molecular Phylogenetics and Evolution* from 2010 – August 2013, 40 of them cited [Hillis and Bull \(1993\)](#) in this manner ([Table S1](#)).

Although rarely cited in this manner in *Molecular Phylogenetics and Evolution* in recent years (but see [Ribeiro et al. \(2012\)](#)), [Hillis and Bull \(1993\)](#) is also still cited as the basis for extrapolating $\geq 70\%$ bootstrap to being an indication of 95% probability of accuracy in empirically focused papers (e.g., [Antibong et al., 2013](#); [Engelbrecht et al., 2013](#); [Guz et al., 2013](#); [Keskin et al., 2013](#); [Kumar et al., 2013](#)). This continues to occur despite [Hillis and Bull's \(1993\)](#) original qualifications as well as those made by other authors on both the interpretation of bootstrap values as a measure of accuracy ([Brown, 1994](#); [Holmes, 2003](#); [Anisimova et al., 2011](#)) and the cases wherein $\geq 70\%$ bootstrap does not equate to 95% probability of accuracy in empirical analyses ([Felsenstein and Kishino, 1993](#); [Soltis and Soltis, 2003](#); [Huelsenbeck and Rannala, 2004](#)).

The standard bootstrap is time-consuming and faster alternatives that have been proposed include rapid bootstrapping (RBS; [Stamatakis et al., 2008](#)), the ultrafast bootstrap approximation (UFBoot; [Minh et al., 2013](#)), and the approximate likelihood-ratio test (aLRT; [Anisimova and Gascuel, 2006](#)). Rapid bootstrapping, as implemented in RAxML ([Stamatakis et al., 2005](#)), is generally applied with the CAT-based approximation of rate heterogeneity (e.g., GTRCAT). Rapid bootstrapping with the GTRCAT model has been criticized as providing inflated bootstrap values, which are likely caused by biased starting trees and low quality tree searches ([Siddall, 2010](#); [Anisimova et al., 2011](#); [Simmons and Norton, 2013](#)). Nonetheless, these methods are not only used as a computational shortcut to regular bootstrapping for supermatrices with several hundred terminals wherein drastic heuristic shortcuts are necessary for likelihood-based analyses (e.g., [Hinchliff and Roalson, 2013](#); [Soltis et al., 2013](#)) – they are also periodically applied to much smaller matrices (e.g., [Colston et al., 2013](#); [Miner et al., 2013](#); [Olsson et al., 2013](#)).

[Minh et al. \(2013\)](#) introduced UFBoot as an alternative to standard bootstrapping and rapid bootstrapping to quickly quantify branch support in likelihood-based analyses of large matrices.

UFBoot applies resampling-estimated log-likelihoods ([Hasegawa and Kishino, 1994](#)) wherein likelihood scores for each character from the original matrix are used to estimate likelihood scores for the bootstrap pseudoreplicates. Important quartet puzzling with NNI branch swapping ([Vinh and von Haeseler, 2004](#)) is used to explore tree space for each bootstrap pseudoreplicate. [Minh et al. \(2013, p. 1189\)](#) reported that, unlike the conservatively biased standard bootstrap, “UFBoot is unbiased for support values higher than 70%” such that “... a split with support of 95% will have a probability of 0.95 to be correct.” They implemented UFBoot, together with standard bootstrapping and the SH-like aLRT, in the program IQ-TREE.

[Anisimova and Gascuel \(2006\)](#) introduced the aLRT as a faster alternative to the bootstrap to quantify branch support. The aLRT tests whether branches have a positive length as opposed to being zero-length and makes nearest-neighbor-interchange (NNI) comparisons at every internal branch in the optimal tree. The aLRT was later modified by [Guindon et al. \(2010\)](#) into the SH-like aLRT (named after [Shimodaira and Hasegawa \(1999\)](#)) by changing the null hypothesis to all three NNI resolutions of a given internal branch being equally likely. [Guindon et al. \(2010\)](#) reported that the SH-like aLRT may perform particularly well relative to the bootstrap when applied to polytomies and very short branches such that the support values assigned to these branches are very low or even zero.

[Anisimova and Gascuel \(2006, p. 550\)](#) asserted that, “The main advantage of the aLRT is that it is much faster than either the [maximum likelihood] bootstrap or the Bayesian inference.” Indeed, the SH-like aLRT has been applied to a very large supermatrix for which the stand likelihood bootstrap or Bayesian MCMCMC ([Yang and Rannala, 1997](#)) inference may not have been computationally tractable ([Pyron et al., 2013](#)).

The simplest way to quantify and compare branch-support values is to restrict phylogenetic analyses to 4-taxon statements. [Suzuki et al. \(2002\)](#) used simulated unrooted 4-taxon statements to test for the fraction of false positives (i.e., type I errors) generated by Bayesian MCMCMC, neighbor joining, and parsimony methods of quantifying branch support using posterior probabilities or the bootstrap. In doing so they generated matrices of 15,000 characters using simple nucleotide-substitution models, and all terminal branches of short and equal length. Taken together, these three factors facilitate accurate phylogenetic inference. Two main types of matrices were generated: those with a zero-length internal branch (hereafter a “star matrix”) and those with 5000 characters generated from all three alternative topologies and then concatenated to create the matrix of 15,000 characters (hereafter a “conflict matrix”). Any resolution generated by one of the phylogenetic-inference methods for either matrix is based on stochastic character variation. Hence high branch-support values should be regarded as false-positives. [Suzuki et al. \(2002\)](#) found that Bayesian posterior probabilities, but not neighbor-joining or parsimony bootstrap values, have a high false-positive rate – even when the true nucleotide-substitution model was used for phylogenetic inference. This behavior of Bayesian posterior probabilities has been attributed to the “star-tree paradox” wherein arbitrary resolutions of polytomies are frequently assigned high posterior probabilities by matrices that include numerous characters because of how the priors are set ([Lewis et al., 2005](#); [Yang, 2007a](#)).

In this study we applied the simple 4-taxon simulations used by [Suzuki et al. \(2002\)](#) for detection of false positive branch support to four novel programs (GARLI ([Zwickl, 2006](#)), IQ-TREE, PhyML ([Guindon and Gascuel, 2003](#)), RAxML) and three methods of quantifying branch support (rapid bootstrapping, UFBoot, aLRT) that have been introduced since 2002. All of these programs and methods represent speed-ups relative to standard likelihood-based bootstrapping as implemented in PAUP* ([Swofford, 2001](#)). The

Download English Version:

<https://daneshyari.com/en/article/5919517>

Download Persian Version:

<https://daneshyari.com/article/5919517>

[Daneshyari.com](https://daneshyari.com)