

RESEARCH ARTICLE

Chronogram or phylogram for ancestral state estimation? Model-fit statistics indicate the branch lengths underlying a binary character's evolution

Jeremy D. Wilson^{1,2}  | Nicolás Mongiardino Koch^{3,4}  | Martín J. Ramírez² ¹Biodiversity and Geosciences Program, Queensland Museum, South Brisbane, Qld, Australia²Museo Argentino de Ciencias Naturales, Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina³Department of Earth & Planetary Sciences, Yale University, New Haven, CT, USA⁴Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA**Correspondence**

Jeremy D. Wilson

Email: jeremydwilson91@gmail.com**Funding information**

Australian Biological Resources Study (ABRS) Taxonomy Research Grant, Grant/Award Number: RG18-03; Fondo para la Investigación Científica y Tecnológica, Grant/Award Number: PICT-2017-2689; National Science Foundation, Grant/Award Number: DEB-2036186; Postdoctoral Fellowship from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina; Yale University Fellowship

Handling Editor: Simone Blomberg**Abstract**

1. Modern methods of ancestral state estimation (ASE) incorporate branch length information, and it has been demonstrated that ASEs are more accurate when conducted on the branch lengths most correlated with a character's evolution; however, a reliable method for choosing between alternate branch length sets for discrete characters has not yet been proposed.
2. In this study, we simulate paired chronograms and phylograms, and generate binary characters that evolve in correlation with one of these. We then investigate (a) the effect of alternate branch lengths on ASE error and (b) whether phylogenetic signal statistics and/or model-fit statistic can be used to select the branch lengths most correlated with a binary character.
3. In agreement with previous studies, we find that ASEs are more accurate when conducted on the branch lengths most correlated with the character. Phylogenetic signal statistics show limited utility for selecting the correct branch lengths, but model-fit statistics are found to be more accurate, with the correct branch lengths generally returning greater model-fit (lower AICc and BIC values). Using this method to choose between alternate branch length sets is more accurate when tree and character properties are more favourable for model optimization, and when shape differences between alternate phylogenies are greater.
4. Our results indicate that researchers conducting ASEs on discrete characters should carefully consider which branch lengths are appropriate, and, in the absence of other evidence, we suggest estimating model-fit values over alternate branch length sets and evolutionary models and choosing the branch length/model combination that returns better model-fit.

KEYWORDS

Akaike information criterion, ancestral state estimation, ancestral state reconstruction, Bayesian information criterion, branch lengths, chronogram, phylogenetic signal, phylogram

1 | INTRODUCTION

Ancestral state estimation (ASE), or ancestral state reconstruction, is the process of estimating the evolutionary history of a character on a phylogeny (Cunningham et al., 1998; Donoghue, 1989; Swofford & Maddison, 1987). This process has been used to reveal the evolution of key innovations across the tree of life, such as the evolution of the orb web in spiders (Kallal et al., 2020) and of reproductive mode in squamates (Pyron & Burbrink, 2014), and to explore the early history of major clades such as flowering plants (Sauquet et al., 2017) and eukaryotes (Skejo et al., 2021). In fact, much of what we know about the history and evolutionary dynamics of morphology, ecology and biogeography across deep time-scales rely on ASE algorithms.

The earliest ASE methods used Maximum Parsimony, reconstructing ancestral states by minimizing state changes over the tree without taking branch lengths into consideration (Maddison, 1991; Swofford & Maddison, 1987). Newer methods based on Maximum Likelihood or Bayesian Inference use models of evolution that incorporate branch lengths, and can account for rate heterogeneity, speciation and extinction rates, and phylogenetic uncertainty (Beaulieu et al., 2013; Huelsenbeck & Bollback, 2001; Maddison et al., 2007; Pagel, 1999b; Pagel et al., 2004).

A remaining issue that has received little attention is how to choose between alternate branch-length sets when conducting ASE using model-based methods. For instance, whether to use a phylogram—a phylogeny on which branch lengths represent amount of evolutionary change, or a chronogram—a phylogeny on which branch lengths represent time (Cascini et al., 2019; Cusimano & Renner, 2014; Litsios & Salamin, 2012). Chronograms are usually used for ASE, perhaps because of an a priori expectation that the probability of change in a given character (e.g. a morphological trait) would depend on the amount of time elapsed. However, studies have shown that rates of morphological change can also strongly correlate with rates of molecular change (Seligmann, 2010), and examples now exist of characters for which, via comparison with secondary evidence, we know that ASE performed on a phylogram is more accurate (e.g. Cascini et al., 2019). Choice of branch lengths for ASE, therefore, remains a broadly relevant problem.

Only a handful of studies have explored this issue. Litsios and Salamin (2012) first demonstrated the sensitivity of ASEs to branch length choice using simulations. They simulated paired chronograms and phylograms, evolved continuous characters on one of them, and looked at the accuracy of ASEs conducted on each branch length set. They found that ASEs were indeed more accurate when performed on the branch length set underlying a character's evolution. They also found a positive correlation between phylogenetic signal, estimated using Blomberg's K (Blomberg et al., 2003) or Pagel's λ (Pagel, 1999a), and reconstruction accuracy, and therefore proposed that researchers should use the branch length set that returns the highest phylogenetic signal for ASE.

Cusimano and Renner (2014) then explored the effect of using alternate branch length sets for ASE on discrete characters, using plant chromosome number datasets. They also found that the choice

of branch lengths could strongly influence the results of ASE but did not find evidence that the phylogenetic signal statistic λ (which can also be applied to discrete characters although this remains controversial; see Harmon, 2018) had any utility for choosing between alternative branch lengths on discrete characters. Although their findings were based on few datasets, they highlighted the need for further investigation into choice of branch lengths for ASE on discrete characters.

In this study, we address this issue by conducting a simulation study to (a) explore the effect of branch length choice on the accuracy of ASEs on discrete, binary characters and (b) assess in detail the utility of several potential test statistics for identifying the branch lengths underlying a discrete character's evolution. We test the utility of three phylogenetic signal statistics: Pagel's λ , Fritz' D (Fritz & Purvis, 2010) and Borges' δ (Borges et al., 2019); the latter two of which have not been investigated previously in this context. We also explore an alternative: using model-fit statistics to choose the branch length set for ASE. These statistics are already used in comparative phylogenetics to compare macroevolutionary models incorporating branch length transformations (Pagel, 1999a), although in this case models incorporating these transformations incur the penalty of an added parameter during model-fitting. Although this penalty cannot be incurred when comparing alternative branch length sets, we hypothesized that model-fit statistics may none-the-less indicate the branch lengths that are more closely correlated with a character's evolution, because these branch lengths would result in higher model-fit values. The model-fit statistics we explore are AICc (Akaike, 1974; Hurvich & Tsai, 1989) and BIC (Schwarz, 1978). Because these model-fit statistics show utility for choosing the correct branch lengths, we further investigate their relationship with ASE error, and identify tree- and character-based properties that affect their utility for choosing the underlying branch lengths of a character.

2 | MATERIALS AND METHODS

We conducted this study using the R statistical language (R Core Team, 2019), in RStudio (RStudio Team, 2019). We made extensive use of the *tidyverse* (Wickham et al., 2019), and phylogenetics packages *APE* (Paradis & Schliep, 2019; Paradis et al., 2004), *CASTOR* (Louca & Doebeli, 2018), *GEIGER* (Harmon et al., 2008), *FOSSILSIM* (Barido-Sottani et al., 2019), *PHANGORN* (Schliep, 2011) and *PHYTOOLS* (Revell, 2012). We also used *GRIDEXTRA* (Aguie, 2017) to construct figures. Additional packages are listed at the beginning of each section.

2.1 | Simulation of chronograms and phylograms

Additional packages: *TREE-SIMGM* (Hagen & Stadler, 2018).

We generated 5,000 ultrametric trees, representing the 'chronograms' for each replicate. Speciation and extinction rates were modelled as age-dependent processes using a Weibull distribution

Maximum Likelihood (Pagel, 1999b). On both branch length sets, we performed ASEs using two Markov models: one with equal transition rates (equal rates), and one with different transition rates in each direction (different rates). AICc and BIC were estimated for all ASEs, leading to four estimates of each statistic per character: one for each Markov model, on both branch length sets.

Next, we estimated the phylogenetic signal statistics Pagel's λ , Fritz's D and Borges' δ on both the chronogram and phylogram, for all characters. Estimation of λ and δ required the choice of an underlying model of evolution because λ is optimized using an underlying model, and δ is estimated based on ancestral state inferences generated using an underlying model. For consistency, we used a Markov model with different transition rates in each direction (i.e. the model used for character simulation) as the underlying model for all estimates of these statistics.

2.4 | Analysis of ASE accuracy and branch length choice

Additional packages: RANDOMFOREST (Liaw & Wiener, 2002).

All analyses were performed separately on the Markov, Hidden Rates and Amplified Hidden Rates character sets, unless otherwise stated.

We first assessed whether ASEs done on the branch lengths that a character evolved on (the 'correct' branch lengths) were more accurate. For each replicate, we identified the Markov model (equal rates or different rates) with the lowest AICc value for both the chronogram and phylogram. For each of these, we calculated *average node error* of the resulting ASE by summing the probabilities of the incorrect state over all nodes in the phylogeny and then dividing by the total number of nodes. Because questions involving ASE often hinge on large probability changes at a few key nodes, we also compared the average error specifically on the five nodes whose state probabilities changed the most between branch length sets, calculating *swing node error*. We then tested whether *average node error* and *swing node error* were lower on the correct branch lengths.

While these analyses explored the accuracy of ASEs across the entire tree, we further explored results obtained at the level of individual nodes, using the Markov character set only. First, we extracted the estimates of ancestral states for all nodes across all replicates, under the same conditions used to simulate the data (i.e. a different-rates Markov model and the correct underlying branch lengths). Ancestral state probabilities were compared through subtraction to (a) the known (i.e. true) ancestral states, to obtain a proxy for accuracy and (b) a scenario in which both states are equiprobable, as a proxy for precision. Then, the difference between the ancestral state probabilities of each node under alternative branch lengths (chronogram and phylogram) was used as a proxy for the sensitivity of ASE to branch length choice. Potential determinants of these three response variables (node-based ASE accuracy, precision and sensitivity) were explored using random forests, which

are relatively robust to multicollinearity, lack of independence and deviations from normality. In all, 10 predictors were calculated for each node, including metrics of depth, number of descendants, relative lengths of the surrounding branches, and the degree to which these changed between the chronogram and phylogram. For brevity, these are defined in Table S1. Regression forests for each response variable were built using 1,000 trees, each of which used a random sample of 1% of total observations (i.e. nodes). The relative importance of each predictor was estimated as the decrease in residual sum of squares introduced by splitting based on that variable (i.e. the increase in node purity), averaged across all trees.

2.5 | Analysis of test statistics for choosing between alternate branch length sets for ASE

Additional packages: ADEPHYLO (Jombart et al., 2010), MGSUB (Ewing, 2020), PERFORMANCEANALYTICS (Peterson & Carl, 2020), PSYCH (Revelle, 2020), RESHAPE2 (Wickham, 2007).

Next, we tested whether values of λ , D , δ , AICc or BIC could be used to identify the branch lengths underlying a character's evolution. For the phylogenetic signal statistics λ , D and δ , we tested whether the correct branch lengths led to higher estimates of phylogenetic signal (corresponding to lower values of D and higher values of λ and δ). For AICc and BIC, we compared the value of these statistics from all four ASEs performed on each character (i.e. from ASEs using both the chronogram and the phylogram, and both the equal-rates and different-rates Markov models) and tested whether the lowest value was estimated on the correct branch lengths. For all five test statistics, we then tested whether they correctly identified underlying chronograms and phylograms at the same level of accuracy, to identify any biases associated with statistics.

Because AICc and BIC showed the highest utility for selecting the correct branch lengths, and because these two statistics returned almost identical results (see Section 3), we conducted several exploratory analyses on AICc to investigate the behaviour of model-fit statistics more generally. First, we tested for a correlation between change in AICc and change in *average node error*, to see whether larger reductions in AICc score corresponded to larger reductions in ASE error. Then, for the Markov character set only, we investigated the influence of several tree- and character-based properties (denoted X1 through X7, see Table 1) on the utility of AICc for selecting the correct branch lengths, using regression models. We tested several predictors that were expected to affect the sensitivity and/or accuracy of model optimization during ASE (X1–6), with the assumption that identifying the correct underlying branch lengths using model-fit statistics would be easier the more effectively the model underlying ASE can be optimized. The final predictor (X7) measured the overall difference in tree shape between the chronogram and the phylogram, with the assumption that the correct branch lengths would be easier to determine when alternate branch length sets were more different. Several predictors (X2–4) were measured on the chronograms, as a proxy for comparison

TABLE 1 Predictor variables included in the regression analysis to assess their effect on the utility of AICc to correctly identify the underlying branch lengths of a character

Property	Definition
X1: Tree size	Number of tips in tree (10–1,000)
X2: Tree depth (chronogram)	Total depth of chronogram (1–100)
X3: Average branch length (chronogram)	Average branch length of chronogram after rescaling to a total depth of 1
X4: Branch length homogeneity (chronogram)	Standard deviation in branch length of chronogram after rescaling to a total depth of 1
X5: Evolutionary rate asymmetry	Calculated as: $ rate\ 1/(rate\ 1 + rate\ 2) - rate\ 2/(rate\ 1 + rate\ 2) $
X6: Proportion of tips in rare state	Number of tips in rare state/total number of tips
X7: Difference in tree shape (between chronogram and phylogram)	Branch score distance (Kuhner & Felsenstein, 1994) between the phylogram and chronogram, after both have been rescaled to have a mean root-to-tip patristic distance of 1 so that this is a measure of the overall difference in branch length proportions (i.e. difference in tree shape)

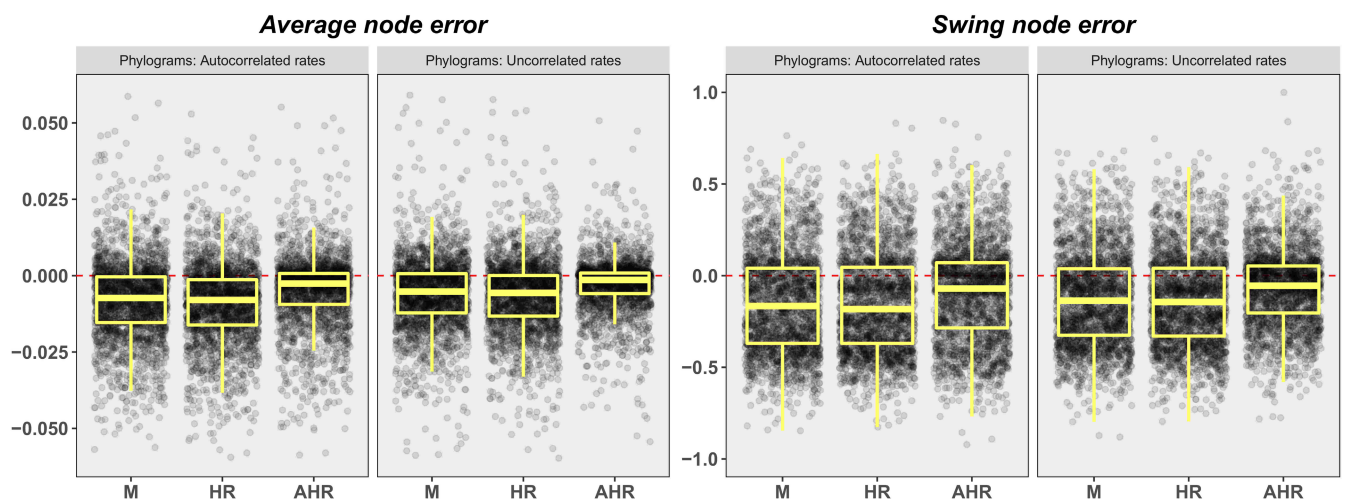


FIGURE 2 Box and whisker plots overlaid on scatter plots showing the change in *average node error* (left) and *swing node error* (right) when ASEs are conducted on the correct versus incorrect branch lengths. The three boxplots in each facet represent characters simulated using Markov (M), hidden rates (HR) and amplified hidden rates (AHR) models. Most values fall below zero, indicating a decrease in error when ASE is conducted on the branch lengths on which a character evolved. To improve interpretability, the y-axis for *average node error* was cut to ± 0.06 , although a minority of replicates ($<1\%$) fall outside this

across replicates—although the exact values of the corresponding phylograms may differ slightly, the overall pattern was expected to be the same. We first tested for correlation among these predictors and removed some to reduce multicollinearity. We then assessed their influence on the response variable using a stepwise multiple logistic regression analysis.

3 | RESULTS

3.1 | Analysis of ASE accuracy and branch length choice

For all character sets, ASEs conducted on the correct branch lengths had lower *average node error* and *swing node error* (Paired Wilcoxon: both $p < 2.2e-16$). *Average node error* was lower on the correct branch lengths in 66%–76% of replicates and *swing node*

error in 63%–71% (Figure 2). On average, *average node error* decreased on the correct branch lengths by 0.4%–0.8%, and *swing node error* by 9%–14% (Figure 2). Results were similar independent of whether evolutionary rates were assumed to be autocorrelated or uncorrelated (Figure 2).

Random forest models were able to explain a substantial fraction of overall variability in both the precision and accuracy of ASEs at individual nodes (33.3% and 25.0%, respectively; Figure 3). Both response variables were highly positively correlated (Spearman's rank correlation: $\rho = 0.84$), showing that strong support for incorrect ASEs is rare in our simulations (e.g. only 3.1% of nodes showed probabilities >0.75 for the incorrect state). In accordance, both variables showed similar patterns of dependence with the predictor set. The most important predictors were all estimates of the relative depth of nodes (including measures of the patristic distance of nodes to the root or the tips descending from it; Figure 3), with both the precision and accuracy of ASEs decreasing with node depth (Figure S1). On

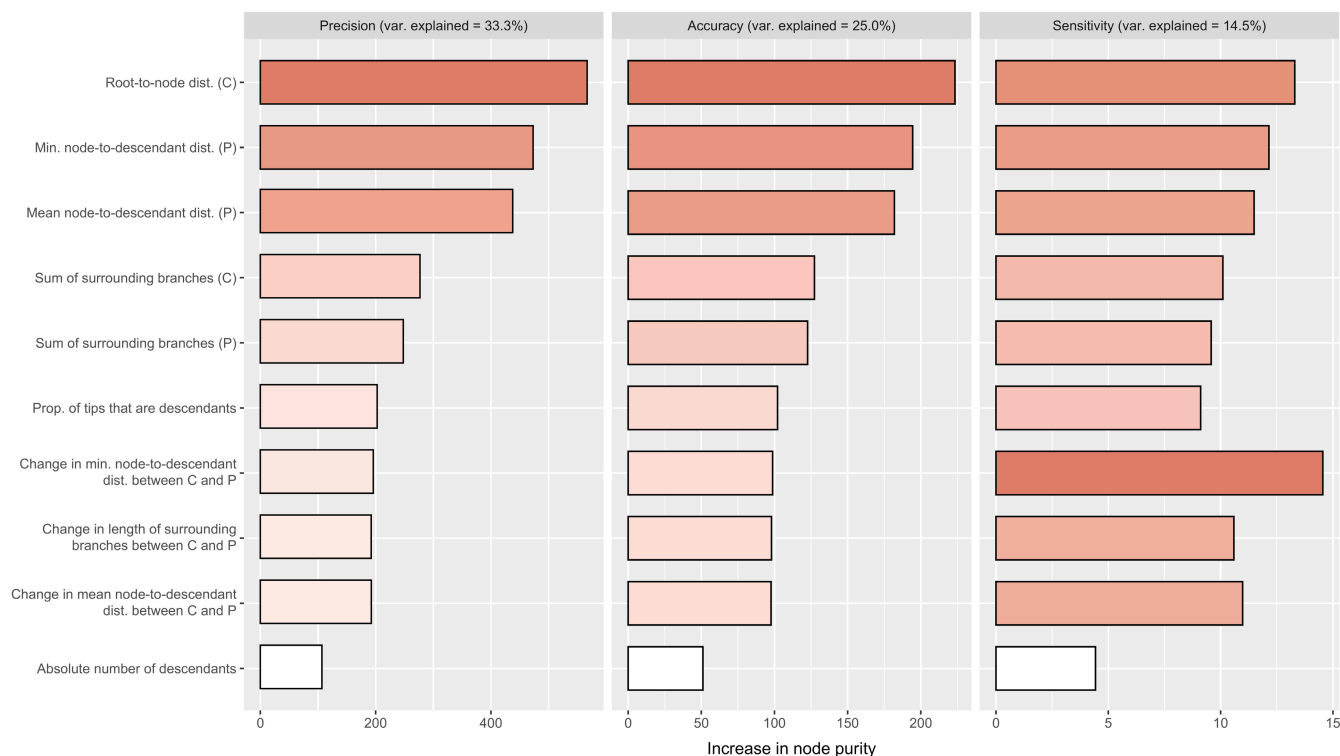


FIGURE 3 Factors determining the precision, accuracy and sensitivity of ASEs at individual nodes. Predictors are ordered according to their importance in explaining node-based precision, estimated as the average increase in node purity in the random forest. Predictors are defined in Table S1; C = chronogram, P = phylogram. The effects of some selected predictors are shown in Figure S1

the other hand, nodes experiencing high levels of sensitivity to the choice of branch length set were relatively rare. Only 0.17% of nodes experienced a change in state probabilities larger than 0.5 when performing ASE in alternative branch length sets. Nonetheless, one in every four topologies contained at least one such highly sensitive node. Node-based ASE sensitivity was not as easy to predict, and our model explained only 14.5% of total variance. While node depth still played a role in determining the sensitivity of ASEs to the use of alternative branch length sets (with shallower nodes being slightly more sensitive to this choice; Figure S1), the most important predictor measured how the minimum patristic distance between a node and its descendants changed between a phylogram and its corresponding chronogram (Figure 3). A potential reason why sensitivity was harder to predict is that highly sensitive nodes seem to generally form clusters (see examples in Figures S2–S4). In these, several consecutive nodes flip their favoured state, in a kind of runaway effect, depending on whether the chronogram or phylogram is used. None of our predictors incorporated information relevant to capture this pattern of node interdependency.

3.2 | Analysis of test statistics for choosing between alternate branch length sets for ASE

Of the statistics tested for their utility in identifying the branch lengths underlying a character's evolution, almost all returned

a significant positive result (Table S2). However, the model-fit statistics AICc and BIC were the most accurate across all character sets, identifying the correct branch lengths about 80% of the time on both the Markov and Hidden Rates character sets (Figure 4). On the Amplified Hidden Rates character set, the accuracy of both model-fit statistics dropped to about 63%, while still outperforming phylogenetic signal statistics. The two model-fit statistics returned almost identical results (Figure 4), selecting the same branch length set in >99% of replicates. Of the phylogenetic signal statistics, D performed the worst, identifying the correct branch lengths close to 50% of the time across character sets (Figure 4). The statistic δ was slightly more accurate at about 60%. The statistic λ was unusual in returning the same value in a large proportion of replicates (almost 40% of replicates for the Markov and Hidden rates character sets, and almost 20% for the Amplified Hidden Rates character set), precluding choice between them. However, for those replicates in which a different value was returned for each branch length set, λ was about 65% accurate for the Markov and Hidden Rates character sets, and 56% accurate for the Amplified Hidden Rates set (Figure 4).

Almost all statistics showed slight differences or 'biases' in their ability to detect underlying chronograms and phylograms; however, the scale of this bias was vastly different (Table S3). The statistic λ was the least biased, only returning statistical significance on the Markov dataset, in which case it correctly identified underlying chronograms slightly more often (6% more) than phylograms. Bias

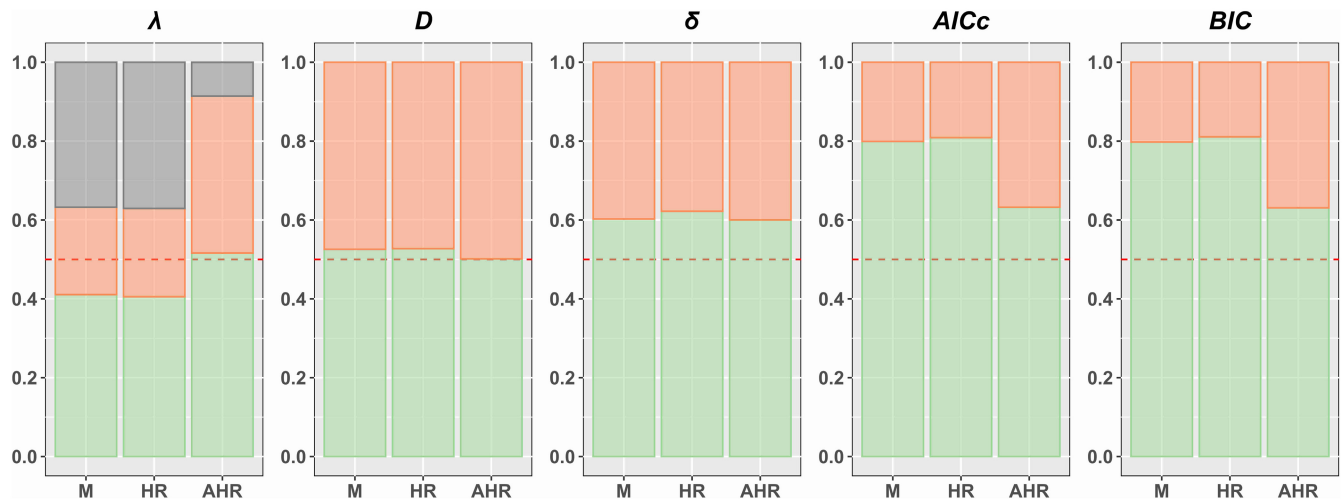


FIGURE 4 Bar chart showing the accuracy of the five test statistics at identifying the correct underlying branch lengths of characters in each of the three sets; the Markov set (M), the hidden rates set (HR) and the amplified hidden rates set (AHR). Green represents correct identifications, red incorrect and grey represents replicates in which both branch length sets returned equal values, precluding a choice. The model-fit statistics (AICc and BIC) were more accurate than the phylogenetic signal statistics (λ , D and δ) across all three character sets, although accuracy of the model-fit statistics dropped markedly on the amplified hidden rates character set, for which model misspecification was the highest

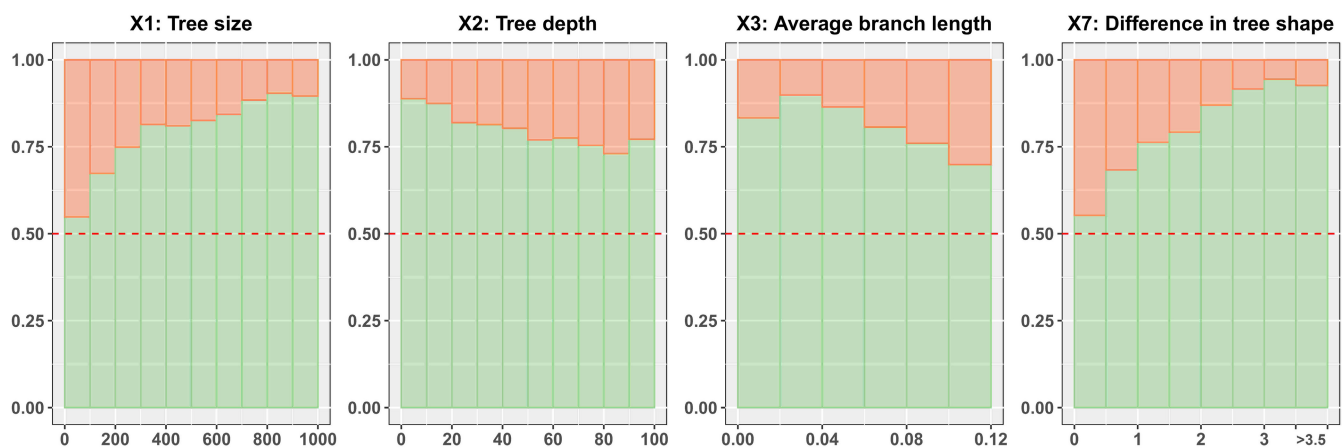


FIGURE 5 Effects of tree- and character-based properties on the utility of AICc for choosing the correct branch lengths. Accuracy increases on larger phylogenies (X1) and when the relative branch lengths of alternative trees are more dissimilar (X7). Accuracy decreases as the total depth and/or the average branch length of the tree increases (X2 and X3, respectively)

in the other two phylogenetic signal statistics was at a greater scale, and consistent across character sets: D correctly identified underlying chronograms about 14%–25% more often than phylograms, and δ identified underlying phylograms 30%–45% more often than chronograms (Table S3). Both model-fit statistics (AICc and BIC) identified underlying phylograms 2%–4% more often on the Markov and Hidden Rates character sets, and underlying chronograms 2% more often on the Amplified Hidden Rates set, which was not found to be significant (Table S3).

The analysis investigating correlation between changes in AICc value and changes in error between ASEs conducted on alternate branch lengths found a significant positive correlation on all character sets (Spearman's rank correlation: all $p < 2.2 \times 10^{-16}$, $\rho = 0.29_{\text{Markov}}, 0.28_{\text{Hidden Rates}}, 0.34_{\text{Amp. Hidden Rates}}$), meaning that

larger improvements in model-fit (i.e. reductions in AICc score) are correlated with larger reductions in ASE error, when comparing alternate branch length sets (Figure S5).

In our analysis investigating the influence of tree- and character-based properties on the utility of AICc for identifying the correct branch lengths, we identified significant and strong correlation between predictors X3: *Average branch length* and X4: *Branch length homogeneity*, and between X5: *Evolutionary rate asymmetry* and X6: *Proportion of tips in rare state* (Figure S6). Due to these strong correlations, we removed X4 and X5 from the analysis in favour of X3 and X6, to avoid excessive multicollinearity, and because these predictors are easier to interpret. Other correlations were identified but were not considered sufficiently strong to warrant the a priori removal of more predictors (Figure S6).

The stepwise procedure identified the model including X1: *Tree size*, X2: *Tree depth*, X3: *Average branch length*, and X7: *Difference in tree shape* as the one with the best fit. All main effects in this model were highly significant ($p \leq 0.00012$; see Table S4), with X1 and X7 having positive effects and X2 and X3 having negative effects (Figure 5; Table S4). A variance inflation factor of 2.51 was returned for X1: *Tree size* (Table S4), indicating that an interaction/correlation exists between this predictor and others in the model (X3 and X7, as indicated by the correlation analysis: Figure S6); however, this level of interaction is generally considered tolerable. Overall, the results of this analysis indicate that the utility of model-fit statistics for identifying the branch lengths underlying a character improves with the number of terminals in the phylogeny and with greater differences in shape (relative branch proportions) between alternative branch length sets. However, its utility decreases as overall tree depth and/or average branch length increase (relative to character transition rates, see discussion) (Figure 5).

4 | DISCUSSION

The core finding of this study is that ancestral state estimations (ASEs) of discrete characters are more accurate when conducted on the branch length set most correlated with the character's evolution, reinforcing previous findings for continuous characters (Litsios & Salamin, 2012), and empirical datasets of discrete characters (Cusimano & Renner, 2014). The importance of branch length choice for ASE is, although slowly, gaining recognition in comparative phylogenetics (e.g. see Cascini et al., 2019; Ramírez et al., 2021), although chronograms are still used in most cases. While a direct association between the amount of elapsed time and the probability of change in a character is sensible, there are scenarios where this link might weaken. While some lineages show a pattern of morphological stability spanning hundreds of millions of years (Cavin & Guinot, 2014; Herrera-Flores et al., 2017) others have experienced burst of morphological change in relatively short timeframes (Hopkins & Smith, 2015; Ronco et al., 2021). Such high levels of rate heterogeneity across lineages can weaken the temporal dependency of character evolution. The history of certain traits can also cause (or be the consequence of) differences in rates of molecular or morphological evolution through physiological, developmental or functional constraints (e.g. Brougham & Campione, 2020; Cascini et al., 2019; Davies & Savolainen, 2006; Smith & Donoghue, 2008), in which case a stronger link with genetic or phenotypic divergence is to be expected (Seligmann, 2010). In light of this, careful consideration of alternative branch length sets, and some justification for why a particular set is chosen, seems warranted if overall accuracy is paramount in ASE.

At the level of individual nodes, we found that although cases where ASEs flip between alternate states depending on the choice of branch length set are relatively rare, they still affected about a quarter of our simulated topologies. Among these, sensitive nodes often formed clusters, defining entire regions of the topology for which

results are highly dependent on the assumptions made (Figures S2–S4). These regions are more common at shallow depths (as ASEs of deeper nodes tend to have low precision and therefore be relatively insensitive to the assumptions made; Figure S1), but more importantly, these nodes are characterized by having highly asymmetrical branch lengths leading to their descendants, with some descendants experiencing less evolutionary change, and therefore having relatively shorter branches on the phylogram (Figure 3). When employing the phylogram for ASE, the short node-to-tip distances that lead to slow-evolving descendants mean that the states observed for them will have a strong effect on ASEs; however when the chronogram is used for ASE, their influence is reduced as distances leading to all descendants even out. Empirical cases involving heterogeneous evolutionary rates (especially at relatively shallow time scales) are likely to require a more careful consideration of the assumptions made in the process of estimating ancestral states (although see Reyes et al., 2018), and are likely to benefit more from the use of objective criteria for selecting among competing branch length sets, such as the model-fit statistics proposed here.

Our results suggest that in the case of discrete characters, using statistics like AICc and BIC to choose the branch length set that leads to the highest model-fit is a promising method for objective branch length choice. Before this study, only phylogenetic signal statistics had been proposed for this, based on the results of Litsios and Salamin (2012), whose investigation focused on continuous characters. In our analysis, these statistics performed poorly, and in the case of D and δ , showed a strong bias in their ability to detect underlying chronograms or phylograms, respectively. In contrast, model-fit statistics identified the correct branch lengths with higher accuracy (almost 20% more often on the Markov and Hidden Rates character sets) and showed only slight biases in their ability to detect either branch length set (<5% in all cases). Our results indicate that although phylogenetic signal statistics measure the strength of the relationship between a phylogeny and a character state distribution (Münkemüller et al., 2012), this is not always a good indicator of whether branch lengths are 'realistic', in terms of their likelihood of having produced the data. In contrast, model-fit statistics like AICc are directly related to the likelihood of a particular model (which incorporates branch lengths) generating the observed state distribution (Posada & Buckley, 2004), and it is therefore intuitive that model-fit should be better on the branch length set that is most realistic.

We found that model misspecification negatively affected the utility of model-fit statistics for identifying the correct underlying branch lengths of a character, with their accuracy lowest on the Amplified Hidden Rates character set. For ASE in this study, we used only simple Markov models that assume constant transition rates across the phylogeny. In empirical datasets, this assumption is expected to be violated to some extent, especially as larger phylogenies are employed that represent greater time-scales and variability in life history (Beaulieu et al., 2013; King & Lee, 2015). On sufficiently large phylogenies, the issue of model misspecification could be reduced by comparing AICc values over not only standard Markov models, but also more complex models such as hidden

rates (Beaulieu et al., 2013; Boyko & Beaulieu, 2021) or rate shift (Grundler & Rabosky, 2020) models that allow for rate heterogeneity. However, on smaller phylogenies that preclude the use of these complex models, researchers need to be aware of the limitations caused by excessive model misspecification to branch length choice.

The utility of model-fit statistics for correctly identifying the underlying branch lengths of a character is affected by several properties that influence the model-fit procedure during ASE, namely tree size (number of taxa), tree depth and average branch length. The positive correlation with tree size was expected because more taxa (i.e. more data) improves model optimization, and should therefore improve estimations of model-fit statistics like AICc. The negative correlation between AICc utility and both tree depth and average branch length may be due to the relationship of these properties with evolutionary rate in our simulation approach. When simulating characters, we assigned character transition rates randomly, and independently of the phylogeny. This meant that deeper trees, and trees with longer average branch lengths, would experience more state changes per branch, on average. This probably led to higher levels of 'saturation' on these trees, weakening the information content of the observed states, reducing overall model accuracy and consequently reducing the utility of model-fit statistics like AICc. The usefulness of AICc for electing branch lengths was also positively correlated with the difference in shape between the chronogram and phylogram. This is also intuitive because as differences between the alternate branch length sets increase, the likelihood that the incorrect branch length set could fit the data better by chance decreases.

5 | CONCLUSIONS

This study reiterates the importance of branch length choice when conducting ASE and provides, for the first time, a method for electing the branch lengths underlying the evolution of a discrete character. This study was conducted on binary characters, but we see no reason why the utility of the method proposed here would not extend to multi-state characters, with the caveat that larger phylogenies would be required for accurate model optimization due to the extra parameters. The weight of evidence indicates that careful consideration and justification of branch-length choice will improve the accuracy of ancestral state estimation. When no other lines of evidence can be used to choose between alternate branch lengths for ASE of a discrete character, choosing the branch lengths that lead to the highest overall model-fit (i.e. the lowest values of statistics like AICc or BIC) is the most appropriate method currently available.

AUTHORS' CONTRIBUTIONS

All authors were involved in conceptualizing and designing this study; J.D.W. and N.M.K. carried out the analyses; J.D.W. led the manuscript writing with substantial revisions from N.M.K. and M.J.R.

ACKNOWLEDGEMENTS

We thank Thomas Guillaume and an anonymous reviewer for their many suggestions that greatly improved this manuscript. J.D.W. was supported by a Postdoctoral Fellowship from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina and an Australian Biological Resources Study (ABRS) Taxonomy Research Grant (RG18-03); N.M.K. by a Yale University Fellowship and NSF project DEB-2036186; M.J.R. by a FONCYT grant PICT-2017-2689. We thank the Arachnology team at Museo Argentino de Ciencias Naturales, and Guilherme H. F. Azevedo, for providing feedback and ideas during the conception of this study. Ivan L. F. Magalhães also provided some preliminary R scripts.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13872>.

DATA AVAILABILITY STATEMENT

Data and scripts are deposited in the Dryad Digital Repository <https://doi.org/10.5061/dryad.z08kprfk> (Wilson et al., 2022).

ORCID

Jeremy D. Wilson  <https://orcid.org/0000-0002-5984-7674>

Nicolás Mongiardino Koch  <https://orcid.org/0000-0001-6317-5869>

Martín J. Ramírez  <https://orcid.org/0000-0002-0358-0130>

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Auguie, B. (2017). *gridExtra: Miscellaneous functions for 'grid' graphics*. R package version 2.3. Retrieved from <https://CRAN.R-project.org/package=gridExtra>
- Barido-Sottani, J., Pett, W., O'Reilly, J. E., & Warnock, R. C. (2019). FossilSim: An r package for simulating fossil occurrence data under mechanistic models of preservation and recovery. *Methods in Ecology and Evolution*, 10, 835–840. <https://doi.org/10.1111/2041-210X.13170>
- Beaulieu, J. M., O'Meara, B. C., & Donoghue, M. J. (2013). Identifying hidden rate changes in the evolution of a binary morphological character: The evolution of plant habit in campanulid angiosperms. *Systematic Biology*, 62, 725–737. <https://doi.org/10.1093/sysbio/syt034>
- Blomberg, S. P., Garland Jr., T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, 57, 717–745. <https://doi.org/10.1111/j.0014-3820.2003.tb00285.x>
- Borges, R., Machado, J. P., Gomes, C., Rocha, A. P., & Antunes, A. (2019). Measuring phylogenetic signal between categorical traits and phylogenies edited by J. Hancock. *Bioinformatics*, 35, 1862–1869. <https://doi.org/10.1093/bioinformatics/bty800>
- Boyko, J. D., & Beaulieu, J. M. (2021). Generalized hidden Markov models for phylogenetic comparative datasets. *Methods in Ecology and Evolution*, 12, 468–478. <https://doi.org/10.1111/2041-210X.13534>
- Brougham, T., & Campione, N. E. (2020). Body size correlates with discrete-character morphological proxies. *Paleobiology*, 46, 304–319. <https://doi.org/10.1017/pab.2020.23>

- Cascini, M., Mitchell, K. J., Cooper, A., & Phillips, M. J. (2019). Reconstructing the evolution of giant extinct kangaroos: Comparing the utility of DNA, morphology, and total evidence. *Systematic Biology*, 68, 520–537. <https://doi.org/10.1093/sysbio/syy080>
- Cavin, L., & Guinot, G. (2014). Coelacanth as 'almost living fossils'. *Frontiers in Ecology and Evolution*, 2, 49. <https://doi.org/10.3389/fevo.2014.00049>
- Cunningham, C. W., Omland, K. E., & Oakley, T. H. (1998). Reconstructing ancestral character states: A critical reappraisal. *Trends in Ecology & Evolution*, 13, 361–366. [https://doi.org/10.1016/S0169-5347\(98\)01382-2](https://doi.org/10.1016/S0169-5347(98)01382-2)
- Cusimano, N., & Renner, S. S. (2014). Ultrametric trees or phylograms for ancestral state reconstruction: Does it matter? *Taxon*, 63, 721–726. <https://doi.org/10.12705/634.14>
- Davies, T. J., & Savolainen, V. (2006). Neutral theory, phylogenies, and the relationship between phenotypic change and evolutionary rates. *Evolution*, 60, 476–483. <https://doi.org/10.1111/j.0014-3820.2006.tb01129.x>
- Donoghue, M. J. (1989). Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. *Evolution*, 43, 1137–1156. <https://doi.org/10.1111/j.1558-5646.1989.tb02565.x>
- Ewing, M. (2020). *mgsub: Safe, multiple, simultaneous string substitution*. R package version 1.7.2. Retrieved from <https://CRAN.R-project.org/package=mgsub>
- Fritz, S. A., & Purvis, A. (2010). Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, 24, 1042–1051. <https://doi.org/10.1111/j.1523-1739.2010.01455.x>
- Goulet, V., Dutang, C., Maechler, M., Firth, D., Shapira, M., & Stadelmann, M. (2017). *Expn: matrix exponential, log, etc.* R Package Version 0.999-2.
- Grundler, M. C., & Rabosky, D. L. (2020). Macroevolutionary analysis of discrete traits with rate heterogeneity. *BioRxiv*. <https://doi.org/10.1101/2020.01.07.897777>
- Hagen, O., Andermann, T., Quental, T. B., Antonelli, A., & Silvestro, D. (2018). Estimating age-dependent extinction: Contrasting evidence from fossils and phylogenies. *Systematic Biology*, 67, 458–474. <https://doi.org/10.1093/sysbio/syx082>
- Hagen, O., Hartmann, K., Steel, M., & Stadler, T. (2015). Age-dependent speciation can explain the shape of empirical phylogenies. *Systematic Biology*, 64, 432–440. <https://doi.org/10.1093/sysbio/syv001>
- Hagen, O., & Stadler, T. (2018). TreeSimGM: Simulating phylogenetic trees under general bellman-Harris models with lineage-specific shifts of speciation and extinction in R. *Methods in Ecology and Evolution*, 9, 754–760. [10.1111/2041-210X.12917](https://doi.org/10.1111/2041-210X.12917)
- Harmon, L. J. (2018). Phylogenetic comparative methods: Learning from trees. Self-published under a CC-BY-4.0 license. Retrieved from <https://lukejharmon.github.io/pcm/>
- Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2008). GEIGER: Investigating evolutionary radiations. *Bioinformatics*, 24, 129–131. <https://doi.org/10.1093/bioinformatics/btm538>
- Herrera-Flores, J. A., Stubbs, T. L., & Benton, M. J. (2017). Macroevolutionary patterns in Rhynchocephalia: Is the tuatara (*Sphenodon punctatus*) a living fossil? *Palaeontology*, 60, 319–328. <https://doi.org/10.1111/pala.12284>
- Hopkins, M. J., & Smith, A. B. (2015). Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 3758–3763. <https://doi.org/10.1073/pnas.1418153112>
- Huelsenbeck, J. P., & Bollback, J. P. (2001). Empirical and hierarchical Bayesian estimation of ancestral states. *Systematic Biology*, 50, 351–366. <https://doi.org/10.1080/106351501300317978>
- Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76, 297–307. <https://doi.org/10.1093/biomet/76.2.297>
- Jombart, T., Balloux, F., & Dray, S. (2010). Adephylo: New tools for investigating the phylogenetic signal in biological traits. *Bioinformatics*, 26, 1907–1909. <https://doi.org/10.1093/bioinformatics/btq292>
- Kallal, R. J., Kulkarni, S. S., Dimitrov, D., Benavides, L. R., Arnedo, M. A., Giribet, G., & Hormiga, G. (2020). Converging on the orb: Denser taxon sampling elucidates spider phylogeny and new analytical methods support repeated evolution of the orb web. *Cladistics*, 37, 298–316. <https://doi.org/10.1111/cla.12439>
- King, B., & Lee, M. S. (2015). Ancestral state reconstruction, rate heterogeneity, and the evolution of reptile viviparity. *Systematic Biology*, 64, 532–544. <https://doi.org/10.1093/sysbio/syv005>
- Kuhner, M. K., & Felsenstein, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution*, 11, 459–468. <https://doi.org/10.1093/oxfordjournals.molbev.a040126>
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2, 18–22.
- Litsios, G., & Salamin, N. (2012). Effects of phylogenetic signal on ancestral state reconstruction. *Systematic Biology*, 61, 533–538. <https://doi.org/10.1093/sysbio/syr124>
- Louca, S., & Doebeli, M. (2018). Efficient comparative phylogenetics on large trees. *Bioinformatics*, 34, 1053–1055. <https://doi.org/10.1093/bioinformatics/btx701>
- Maddison, W. P. (1991). Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Systematic Biology*, 40, 304–314. <https://doi.org/10.1093/sysbio/40.3.304>
- Maddison, W. P., Midford, P. E., & Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, 56, 701–710. <https://doi.org/10.1080/10635150701607033>
- Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffrers, K., & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, 3, 743–756. <https://doi.org/10.1111/j.2041-210X.2012.00196.x>
- Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., & Pearse, W. (2012). Caper: Comparative analyses of phylogenetics and evolution in R. R Package Version 0.5.2.
- Pagel, M. (1999a). Inferring the historical patterns of biological evolution. *Nature*, 401, 877–884.
- Pagel, M. (1999b). The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology*, 48, 612–622.
- Pagel, M., Meade, A., & Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*, 53, 673–684. <https://doi.org/10.1080/10635150490522232>
- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290. <https://doi.org/10.1093/bioinformatics/btg412>
- Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Peterson, B. G., & Carl, P. (2020). *PerformanceAnalytics: Econometric tools for performance and risk analysis*. R package version 2.0.4. Retrieved from <https://CRAN.R-project.org/package=PerformanceAnalytics>
- Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53, 793–808. <https://doi.org/10.1080/10635150490522304>
- Puttick, M. N. (2016). Partially incorrect fossil data augment analyses of discrete trait evolution in living species. *Biology Letters*, 12, 20160392. <https://doi.org/10.1098/rsbl.2016.0392>
- Pyron, R. A., & Burbrink, F. T. (2014). Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. *Ecology Letters*, 17, 13–21. <https://doi.org/10.1111/ele.12168>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

- Ramírez, M. J., Magalhaes, I. L., Derkarabetian, S., Ledford, J., Griswold, C. E., Wood, H. M., & Hedin, M. (2021). Sequence capture phylogenomics of true spiders reveals convergent evolution of respiratory systems. *Systematic Biology*, 70, 14–20. <https://doi.org/10.1093/sysbio/syaa043>
- Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 2, 217–223. [10.1111/j.2041-210X.2011.00169.x](https://doi.org/10.1111/j.2041-210X.2011.00169.x)
- Revelle, W. (2020). *Psych: Procedures for personality and psychological research*. Northwestern University. Retrieved from <https://CRAN.R-project.org/package=psych>
- Reyes, E., Nadot, S., von Balthazar, M., Schönenberger, J., & Sauquet, H. (2018). Testing the impact of morphological rate heterogeneity on ancestral state reconstruction of five floral traits in angiosperms. *Scientific Reports*, 8, 9473. <https://doi.org/10.1038/s41598-018-27750-1>
- Ronco, F., Matschiner, M., Böhne, A., Boila, A., Büscher, H. H., El Taher, A., Indermaur, A., Malinsky, M., Ricci, V., & Kahmen, A. (2021). Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature*, 589, 76–81. <https://doi.org/10.1038/s41586-020-2930-4>
- RStudio Team. (2019). *RStudio: Integrated development environment for R* (Vol. 770). RStudio, Inc.
- Sauquet, H., von Balthazar, M., Magallón, S., Doyle, J. A., Endress, P. K., Bailes, E. J., Barroso de Morais, E., Bull-Hereñu, K., Carrière, L., Chartier, M., Chomicki, G., Coiro, M., Cornette, R., El Ottra, J. H. L., Epicoco, C., Foster, C. S. P., Jabbour, F., Haevermans, A., Haevermans, T., ... Schönenberger, J. (2017). The ancestral flower of angiosperms and its early diversification. *Nature Communications*, 8, 16047. <https://doi.org/10.1038/ncomms16047>
- Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27, 592–593. <https://doi.org/10.1093/bioinformatics/btq706>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464. <https://doi.org/10.1214/aos/1176344136>
- Seligmann, H. (2010). Positive correlations between molecular and morphological rates of evolution. *Journal of Theoretical Biology*, 264, 799–807. <https://doi.org/10.1016/j.jtbi.2010.03.019>
- Skejo, J., Garg, S. G., Gould, S. B., Hendriksen, M., Tria, F. D. K., Bremer, N., Franjević, D., Blackstone, N. W., & Martin, W. F. (2021). Evidence for a syncytial origin of eukaryotes from ancestral state reconstruction. *Genome Biology and Evolution*, 13, evab096. <https://doi.org/10.1093/gbe/evab096>
- Slater, G. J., Harmon, L. J., & Alfaro, M. E. (2012). Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution*, 66, 3931–3944. <https://doi.org/10.1111/j.1558-5646.2012.01723.x>
- Smith, S. A., & Donoghue, M. J. (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science*, 322, 86–89. <https://doi.org/10.1126/science.1163197>
- Swofford, D. L., & Maddison, W. P. (1987). Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences*, 87, 199–229. [https://doi.org/10.1016/0025-5564\(87\)90074-5](https://doi.org/10.1016/0025-5564(87)90074-5)
- Tao, Q., Tamura, K., Battistuzzi, F. U., & Kumar, S. (2019). A machine learning method for detecting autocorrelation of evolutionary rates in large phylogenies. *Molecular Biology and Evolution*, 36, 811–824. <https://doi.org/10.1093/molbev/msz014>
- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21, 1–20. Retrieved from <http://www.jstatsoft.org/v21/i12/>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., & Hester, J. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4, 1686. <https://doi.org/10.21105/joss.01686>
- Wilson, J. D., Mongiardino Koch, N., & Ramírez, M. J. (2022). Chronogram or phylogram for ancestral state estimation? Model-fit statistics indicate the branch lengths underlying a binary character's evolution: R scripts and simulated trees. *Dryad Digital Repository*. <https://doi.org/10.5061/dryad.z08kprfk>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Wilson, J. D., Mongiardino Koch, N. & Ramírez, M. J. (2022). Chronogram or phylogram for ancestral state estimation? Model-fit statistics indicate the branch lengths underlying a binary character's evolution. *Methods in Ecology and Evolution*, 13, 1679–1689. <https://doi.org/10.1111/2041-210X.13872>