RESEARCH ARTICLE

JOURNAL OF
EXPERIMENTAL
ZOOLOGY
PART B
MOLECULAR AND DEVELOPMENTAL EVOLUTION

# Phylogenetic Uncertainty Can Bias the Number of Evolutionary Transitions Estimated From Ancestral State Reconstruction Methods

SEBASTIAN DUCHÊNE[1,2]*
AND ROBERT LANFEAR[3]*
[1]School of Biological Sciences, University of Sydney, Sydney, Australia
[2]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, Sydney Medical School, University of Sydney, Sydney, Australia
[3]Department of Biological Sciences, Macquarie University, Sydney, Australia

**ABSTRACT** Ancestral state reconstruction (ASR) is a popular method for exploring the evolutionary history of traits that leave little or no trace in the fossil record. For example, it has been used to test hypotheses about the number of evolutionary origins of key life-history traits such as oviparity, or key morphological structures such as wings. Many studies that use ASR have suggested that the number of evolutionary origins of such traits is higher than was previously thought. The scope of such inferences is increasing rapidly, facilitated by the construction of very large phylogenies and life-history databases. In this paper, we use simulations to show that the number of evolutionary origins of a trait tends to be overestimated when the phylogeny is not perfect. In some cases, the estimated number of transitions can be several fold higher than the true value. Furthermore, we show that the bias is not always corrected by standard approaches to account for phylogenetic uncertainty, such as repeating the analysis on a large collection of possible trees. These findings have important implications for studies that seek to estimate the number of origins of a trait, particularly those that use large phylogenies that are associated with considerable uncertainty. We discuss the implications of this bias, and methods to ameliorate it. *J. Exp. Zool. (Mol. Dev. Evol.) 9999B: 1–8, 2015.* © 2015 Wiley Periodicals, Inc.

*J. Exp. Zool.*
*(Mol. Dev. Evol.)*
*9999B:1–8, 2015*

**How to cite this article: Duchêne S, Lanfear R. 2015. Phylogenetic uncertainty can bias the number of evolutionary transitions estimated from ancestral state reconstruction methods. J. Exp. Zool. (Mol. Dev. Evol.) 9999:1–8.**

Reconstructing and explaining species' evolutionary history is one of the central goals of biology. Ancestral state reconstruction (ASR) is a popular approach that combines phylogenies and models of trait evolution to infer how species' traits have changed over evolutionary time. ASR is potentially very useful, because it allows us to study the evolutionary history of traits that leave little or no trace in the fossil record. For example, recent studies have used ASR to understand the evolution of reproductive mode in reptiles (Pyron and Burbrink, 2013), genome size in flowering plants (Beaulieu et al., 2010), salt tolerance, and photosynthetic pathways in grasses (Edwards and Smith, 2010; Bennett et al., 2013; Bromham and Bennett, 2014), placentas in mammals (Elliot

Grant sponsor: Colombian Government; grant sponsor: University of Sydney; grant sponsor: Australian Research Council.

Additional supporting information may be found in the online version of this article at the publisher's web-site.

*Correspondence to: Sebastian Duchêne, Charles Perkins Centre D17, University of Sydney, NSW 2006, Sydney, Australia.
E-mail: sebastian.duchene@sydney.edu.au

*Correspondence to: Robert Lanfear , Department of Biological Sciences, Faculty of Science, Macquarie University, NSW 2109, Sydney, Australia. E-mail: robert.lanfear@mq.edu.au

Received 10 April 2015; Accepted 15 June 2015
DOI: 10.1002/jez.b.22638

and Crespi, 2009), and vocal communication in birds (Odom et al., 2014).

ASR relies on three key elements: accurate trait data, a model of trait evolution, and an estimate of the phylogenetic relationships of the clade of interest. The importance of accurate trait data is widely appreciated and often forms the majority of the work that goes into an ASR study. The model of trait evolution is also vital to the accuracy of the inferences that are made, and is the focus of a huge amount of ongoing work and debate (FitzJohn et al., 2009; Beaulieu et al., 2013; Revell, 2014). In contrast, the accuracy of the phylogenies used in ASR has received less attention (Diaz-Uriarte and Garland, '98; Huelsenbeck and Bollback, 2001). All ASR methods rely on an estimate of the phylogeny, but how good does this estimate have to be? Can uncertainty or inaccuracy in the tree topology itself bias ASR analyses?

A focus on phylogenetic uncertainty is timely because there has been a dramatic increase in the number of ASR projects based on extremely large phylogenies, often covering many or most of the species in a clade. For example, recent years have seen the publication of many large phylogenies that have been used for ASR studies, including trees of: 9,993 birds (Jetz et al., 2012); 32,223 angiosperms (Zanne et al., 2014); 1,230 grasses (Edwards and Smith, 2010); 2,871 amphibians (Pyron and Wiens, 2011); and 7,822 fish (Rabosky et al., 2013). ASR studies that use phylogenies of this size make inferences of broad scope that can have far-reaching implications, so it is important that we understand the limits that phylogenetic accuracy can place on inferences made from ASR studies.

In this study, we focus on an approach in which the phylogeny is estimated independently of the ASR analysis itself. It is possible, and arguably preferable, to perform a joint inference of the phylogeny and the ancestral traits, such that the plausibility of particular ancestral state reconstructions can inform the estimate of the phylogeny (Huelsenbeck and Bollback, 2001; Lemey et al., 2009). However, the joint inference approach remains relatively uncommon, perhaps because it is not computationally feasible for very large datasets. Instead, most ASR studies either use a published phylogeny, or estimate a phylogeny from molecular data before proceeding with reconstructing the evolutionary history of the trait(s) of interest on that phylogeny.

All estimates of phylogenies are uncertain, but how could this uncertainty affect inferences made from ASR? We first provide a somewhat contrived example, and then assess this question more generally by using simulations. For simplicity, we focus throughout on attempts to estimate the number of transitions between two states of a single trait. However, we note that the reasoning we develop is likely to apply to many other uses of ASR.

To exemplify the problem, imagine you are interested in estimating the number of transitions between two states of a single trait, such as viviparous and oviparous reproduction, over the history of a clade. Imagine also that the rate of transition is relatively low, as is often the case for ASR studies, since fast evolving traits leave little phylogenetic signal and are less amenable to study. If you knew the true phylogeny, the true model of evolution, and data on the trait of interest in all extant taxa, you would have a good chance of inferring the correct number of transitions. For example, on the phylogeny shown in Figure 1A all reasonable methods would infer a single change in the trait. But what happens if you don't know the true phylogeny? In the face of phylogenetic uncertainty, you will tend to infer more transitions than actually occurred, simply because there are more ways to intermingle taxa with different trait values than there are ways to keep them separate (Fig. 1B). Thus, in Figure 1B a relatively minor change in the phylogeny leads to the reconstruction of three transitions rather than one. The problem is that among the vast number of possible trees, very few trees give the correct number of transitions, and the vast majority give too many transitions. That is, phylogenetic inaccuracy tends to lead to an overestimation of the true number of transitions that have occurred on a tree.

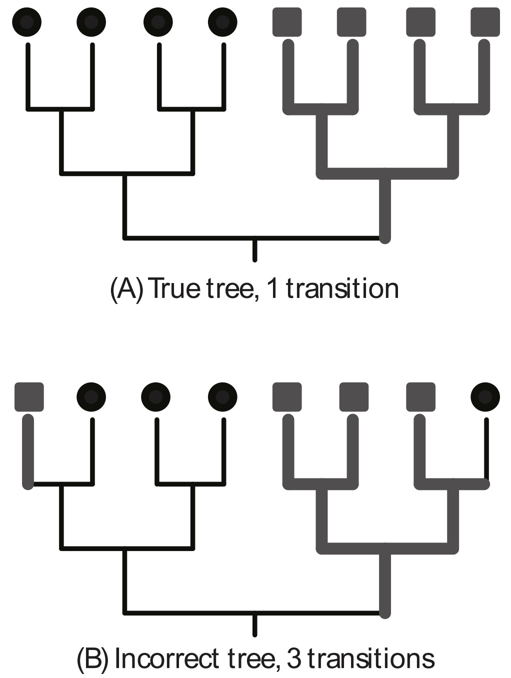

**Figure 1.** Illustration of bias in estimating the number of transitions during ancestral state reconstruction. (A) When the true tree is known, most reconstruction methods would suggest a single transition. (B) When the phylogeny is incorrect, most trees tend to lead to an overestimate of the number of transitions. In this example, a small error in reconstruction in which two species are interchanged leads to the reconstruction of three transitions instead of one.

To demonstrate this, we can calculate the scale of the problem for a simple case study. Imagine that we are interested in reconstructing the number of transitions between two states in a clade of *N* species. We start with a true tree of *N* species that is balanced and rooted, and that is constructed such that all of the species in the left-hand clade (clade 1) share one value of the trait, and all of the species in the right-hand clade (clade 2) share the other value of the trait. Figure 1A provides an example for 8 species. In this case, we only need to infer a single transition to explain the data, although we would not know without additional information the state of the trait at the root of the tree. We can now ask what proportion of all possible trees will overestimate the true number of transitions.

Calculating the proportion of trees that will overestimate the number of transitions is relatively simple. For the case presented here, we will correctly infer a single transition as long as all of the species in either clade 1 and/or clade 2 remain monophyletic. All other possible trees will infer more than one transition. The proportion of trees that overestimate the number of transitions is simply the number of trees that overestimate the number of transitions, divided by the number of possible trees for *N* taxa (see Appendix for derivation):

$$\frac{(2N-3)!! - (N-3)!![(N-3)!! + (2N-4)]}{(2N-3)!!}$$

For the case study here of a single true transition, this shows that in all but trivially small trees, the vast majority of trees will overestimate the true number of transitions. For example, in a study of just 6 taxa, more than 96% of all possible trees will overestimate the number of transitions. This increases rapidly with the number of taxa, such that the proportion is >99.9% for 10 taxa, and >99.999% for 15 taxa. This simple example, although contrived, implies that as long as our estimate of the true phylogeny is uncertain we will often overestimate the number of transitions between states for slowly evolving traits.

But wouldn't such a bias be accounted for by standard methods of accounting for phylogenetic uncertainty? Researchers typically take one of two approaches to phylogenetic uncertainty: they either ignore it or they attempt to account for it by repeating the inference across a large collection of trees. Many of the most recent studies ignore phylogenetic uncertainty, and infer the number of transitions on a single estimate of the "best" tree. This is lamentable, because all phylogenetic estimates are uncertain, and ignoring uncertainty in the tree topology and branch lengths can only lead to overconfidence in other parameter estimates (Huelsenbeck and Bollback, 2001). In these cases, we expect the number of transitions to be overestimated. ASR studies that do account for phylogenetic uncertainty tend to do so by repeating the ASR on a large collection of trees (e.g., 100 bootstrap trees, or 100 trees from the posterior distribution of trees from a Bayesian analysis), to infer a distribution of the parameters of interest. However, although this approach is common, it may not be sufficient to account for the bias we describe: if the overwhelming majority of trees overestimate the number of transitions, then the distribution of the number of transitions calculated from a large collection of trees will also be biased. Indeed, because the proportion of trees that overestimate the number of transitions is so high, it is feasible that even very large samples of trees may all overestimate the true number of transitions.

We use simulations to investigate this problem, by comparing estimates of the number of transitions from ancestral state reconstructions on the true tree to estimates derived from trees reconstructed from simulated DNA sequence data. We show that the effect we describe above can lead to large biases in the estimated number of transitions, and that this problem is often not accounted for by conducting the analyses over a large collection of trees. We suggest that this bias should be considered when interpreting the results of such analyses, and that it is also likely to occur in applications of ASR that we have not studied here, such as the estimation of rates of transition, or of rates of evolution of continuous parameters such as body size.

## MATERIALS AND METHODS

We used a simulation framework to investigate the impact of phylogenetic uncertainty, tree size, and transition rate on the estimation of the number of transitions between two states of a binary trait. In each simulation, we generated 10 replicate datasets, where each replicate involved simulating a phylogeny, simulating a DNA sequence alignment, and simulating binary character data on the true phylogeny. In total we examined 54 simulation conditions, comprising three tree sizes (50, 100, and 500 taxa), six possible alignment lengths (0, 10, 20, 50, 100, and 1000 nucleotides), and three possible transition rates (corresponding to 3, 10, and 30 expected transitions). Thus, in total we simulated 540 datasets. Below, we describe these simulations in more detail, and we then describe how we used the simulations to test for bias in estimating the number of evolutionary origins of a trait.

### Simulation of Phylogenetic Trees

We simulated phylogenetic trees of 50, 100, and 500 taxa with a root-node age of 1 Myr under the Yule speciation process, as implemented in TreeSim v1.9.1 (Stadler, 2011). This implementation of the Yule process produces ultrametric trees with branch lengths in units of time, known as chronograms. We simulated the substitution rate along these chronograms using the lognormal relaxed-clock model (Drummond et al., 2006), implemented in NELSI v0.21 (Ho et al., 2015). We parameterized the clock model using a mean of 0.1 substitutions per site per year (subs/site/year) and a standard deviation of 10% of the mean. We divided the branch lengths of the chronograms by the branch-specific rates to obtain phylograms, in which the branch lengths correspond to the expected number of substitutions per site (subs/site).

### Simulation of Nucleotide Alignments

We simulated nucleotide evolution along the phylograms under the Jukes-Cantor (JC) substitution model (Jukes and Cantor, 1969) to obtain sequence alignments of 0, 10, 20, 50, 100, and 1000 nucleotides. We chose the JC model to avoid the need to select arbitrary parameterizations of more complex models, and because our study does not involve an investigation of substitution model misspecification. We chose the range of alignment lengths after some preliminary analyses, which indicated that this collection of alignment lengths resulted in reconstructed trees with the full range of uncertainty for most datasets. Trees reconstructed with alignments lengths of 0 nucleotides have very low support and high uncertainty, whereas trees estimated from alignments of 1000 nucleotides tend to have much lower uncertainty and higher support.

### Simulation of Trait Data

We simulated stochastic evolution of a two-state character with states A and B along the simulated chronograms, using the *sim. history* function from Phytools v0.4 (Revell, 2012). This method requires a transition rate matrix, known as $Q$. We parameterized $Q$ to produce three different expected numbers of transitions along each of the 180 simulated trees: 3, 10, and 30, which we designate as low, medium, and high, respectively, throughout this study. To calculate the transition probabilities we divided the expected number of transitions by the total tree length. Our parameterization of $Q$ consisted in symmetric transition rates, such that the probability of observing a transition from A to B is the same as the reverse. We set the root-node state to A in all of our simulations. The output from these simulations is the character state for each taxon in the tree.

### Testing for Bias in Estimates of the Number of Transitions

To test for bias in ancestral state reconstructions, we compared the number of transitions estimated on the true tree (i.e., the tree under which the character data were simulated) with the number estimated on trees reconstructed from the simulated sequence data. This approach ensures that the only difference between the two approaches is in the underlying phylogeny used in the ASR, that is whether we use the true tree or an estimate of it. This is important, since even when the true tree and the true model are used, one cannot be guaranteed to recover the true number of transitions, especially if the number of transitions is very large in proportion to the number of taxa in tree. Thus, our comparison automatically accounts for any error or bias inherent in the reconstruction itself, and allows us to focus exclusively on the impact of phylogenetic uncertainty on the estimates of the number of transitions.

We estimated phylogenetic trees from our simulated alignments using a Bayesian Markov chain Monte Carlo (MCMC) sampling method implemented in BEAST v2.1 (Bouckaert et al., 2014). We matched the models in BEAST to those used in our simulations by using the Yule tree prior, the JC substitution model, and the lognormal relaxed-clock model. We calibrated the molecular clock by specifying the age of the root-node in the form of a uniform prior distribution with minimum and maximum bounds of 0.9 and 1.1, respectively. We sampled from the posterior distribution every $10^3$ steps from an MCMC length of $10^7$ steps. We discarded the first 10% steps as burn-in. To assess sufficient sampling from the stationary distribution we verified that the effective sample size for all parameters was at least 200, using CODA v0.16 (Plummer et al., 2006). If the effective sample size for any parameter was less than 200, we increased the chain length by 20% and repeated the analysis.

To estimate the number of transitions along the trees we used stochastic character mapping, SIMMAP (Huelsenbeck et al., 2003), as implemented in Phytools. In this method, the conditional likelihood for each state is calculated at every node, and then simulations are conducted by sampling from the posterior distribution of the character states (Huelsenbeck et al., 2003; Bollback, 2006). For each of our analyses we sampled 200 trees from the posterior distribution, which we refer to as "reconstructed trees." We conducted 100 simulations for the SIMMAP method to estimate the number of transitions for each of the reconstructed trees. We specified symmetric transition rates to match the model used to generate the data, and we calculated the mean estimated number of transitions across the 100 simulations of the SIMMAP method for each tree. However, because our analyses included 200 trees from the posterior, we considered the uncertainty in the estimates as the range of mean values obtained for these trees. For comparison, we also used this method to estimate the number of transitions along the true tree.

We quantified errors in the estimated number of transitions by comparing the estimates obtained with the true tree, with those from the reconstructed trees. We calculated the ratio of the number of transitions estimated on the reconstructed trees to the number of transitions estimated on the true tree. This ratio describes the over- or underestimation of the number of transitions as a result of using reconstructed trees. For example, a ratio of 3 indicates that a reconstructed tree had three times as many transitions as the true tree. A ratio of 0.5 indicates that the reconstructed tree had half as many transitions as the true tree. Thus, if the distribution of the number of transitions calculated from the reconstructed trees overlaps with 1, that distribution includes the value estimated on the true tree.

To investigate the relationship between node support and the estimated number of transitions, we obtained the highest clade credibility (HCC) tree for each simulated data set and calculated the mean posterior probability for all nodes. Note that we do not use the HCC tree to estimate the number of transitions, however, the mean posterior support of this tree serves as a measure of phylogenetic uncertainty in the reconstructed trees. For instance, if the HCC tree for a simulated data set has a mean posterior probability of 0.02, the number of different tree topologies in the posterior is expected to be very large, such that there is high phylogenetic uncertainty. On the contrary, if the posterior probability of the HCC tree is 0.99, then the number of tree topologies in the posterior should be very small, with low phylogenetic uncertainty. The computer code to reproduce our simulations and analyses is freely available in an online repository (http://bit.ly/18JDdkh).

## RESULTS AND DISCUSSION

In this study, we used simulations to ask whether estimates of the number of transitions can be biased by phylogenetic uncertainty. Our results show that phylogenetic uncertainty can lead to severe biases in ancestral state reconstruction, and that conducting the analysis on a large collection of trees is usually insufficient to account for this bias (Fig. 2). In the vast majority of cases the number of transitions was over-estimated, regardless of the simulation conditions. Notably, this bias also applies when the level of phylogenetic uncertainty was within levels commonly considered as acceptable (e.g., mean posterior node support >90%). These results have important implications for the interpretation of the results of studies that seek to reconstruct the number of evolutionary transitions between traits: as long as the phylogenetic tree is not known with certainty, reconstructions tend to overestimate the number of transitions that have occurred.

As expected, the bias in the estimated number of transitions was highest when the expected number of transitions was lowest (Fig. 2). For example, in our simulations with three expected transitions (Fig. 2, top row), the number of transitions on reconstructed trees was frequently more than three-fold higher than the number estimated on the true tree. But with thirty expected transitions (Fig. 2, bottom row), the bias was rarely higher than three-fold. This reduction in bias occurs because smaller numbers of expected transitions result in a larger proportion of trees that overestimate the number of transitions, and because there is an upper limit to the number of transitions that can be reconstructed on a tree of a given size. When considering the impacts of this bias on empirical studies, it is important to note that it is not the absolute number of transitions that is important, but the overall rate of transition. For example, our simulations of 10 transitions on a tree of 50 taxa represent roughly 1 transitions occurring on every 10 branches. Our results show that in this case even highly supported trees (mean posterior support >0.9) can estimate almost twice the true number of transitions.

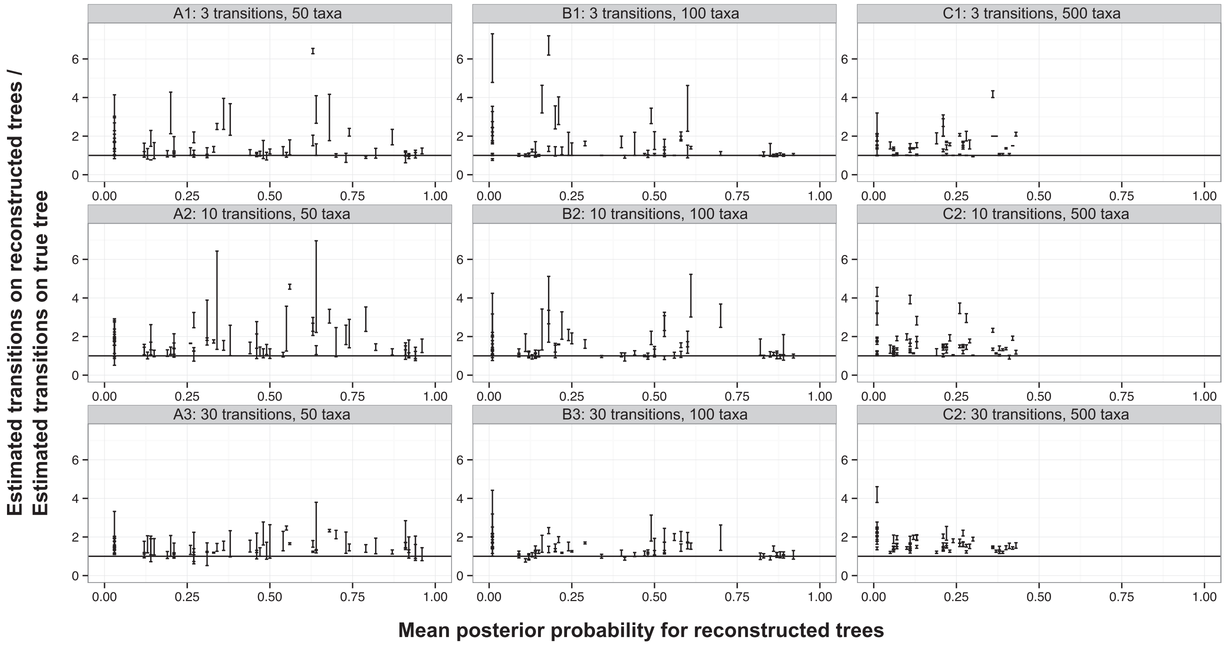

**Figure 2.** Error in the estimated number of transitions vs. the mean posterior probability (node support) for simulations with different numbers of transitions and number of taxa. The *y*-axis is the error measured as the estimated number of transitions on the reconstructed tree divided by the estimated number of transitions on the true tree, such that values higher than one correspond to overestimates. Each bar corresponds to a single simulated data set, and represents the range of estimates for the 200 trees taken from the MCMC, that is it represents an estimate of the range of the posterior distribution of the number of transitions for a given simulation. The horizontal line corresponds to an error value of 1; bars that overlap this value represent simulations in which the number of transitions estimated on the true tree is included in the range of the number of transitions estimated from the posterior distribution of reconstructed trees.

The links between phylogenetic uncertainty and bias in the estimated number of transitions depended on the number of taxa in the simulated tree. For trees of 50 and 100 taxa, higher node support was generally associated with smaller bias (Fig. 2 columns A and B). But for trees of 500 taxa this trend was less marked, perhaps because even our largest simulated alignments of 1000 sites did not provide sufficient information to resolve the 500 taxon trees to more than an average of 0.4 mean posterior node support (Fig. 2 column 3). Nevertheless, even highly supported trees, i.e., when average posterior node support was >90%, often estimated twice the true number of transitions (Fig. 2, columns 1 and 2). This occurs because small portions of the trees have low node support, such that the number of transitions will tend to be overestimated if they occur in these sections of the tree (some examples are provided in Supplementary Fig. S1).

Accounting for uncertainty in the phylogeny by repeating the analysis on 200 trees from the posterior distribution was often insufficient to address the bias. In most of our simulations, the distribution of the number of transitions from the reconstructed trees did not overlap with the number of transitions reconstructed on the true tree (shown on Fig. 2 by the large number of cases in which the distributions do not overlap with the line at a value of 1 on the *y*-axis). This shows that there were frequently cases in which none of the trees in the posterior sample were sufficiently accurate to reconstruct the true number of transitions. Surprisingly, this was also the case for some of our simulations in which the mean posterior node support was greater than 90%. Perhaps most importantly, our results suggest that studies which do not attempt to account for phylogenetic uncertainty are liable to overestimate the number of transitions by a large factor. For example, we frequently observed cases in which some trees from the posterior distribution overestimated the true number of transitions by fourfold, even when the mean posterior node support was quite high (Fig. 2).

### Implications for Empirical ASR Studies

This study presents the results from a limited set of simulations. By design, these simulations cover a large range of phylogenetic uncertainty, most of which would be deemed far too high to be acceptable in empirical ASR studies. Nevertheless, the simulations give a coherent picture that phylogenetic uncertainty of any extent tends to bias estimates of the number of evolutionary transitions between states, even when all other details of the analysis (e.g., the model of trait evolution) are correct. In particular, our results suggest that estimates of the number of transitions from empirical studies are likely to be biased upwards, and that this bias could be particularly severe in studies that rely on a single estimate of the phylogeny. Our results also show that this overestimation cannot always be addressed by simply repeating the ASR on a large collection of trees.

The extent to which empirical studies will overestimate the number of transitions between states will depend on both the accuracy of the phylogeny that is used, and the methods employed to account for uncertainty. Studies which use very well-resolved trees, are unlikely to dramatically overestimate the number of transitions due to the bias we describe, regardless whether or how they attempt to account for uncertainty in the tree topology and branch lengths. However, our results show that even quite well supported trees, with mean posterior node support higher than 90%, can still overestimate of the number of transitions by a factor of two. Empirical studies that rely on large phylogenies of thousands of taxa may be particularly prone to the bias we describe. This is because such studies typically rely on phylogenies estimated from relatively small and sometimes sparse alignments, which can be associated with considerable uncertainty. Furthermore, such studies typically use approaches that rely on a single estimate of the phylogeny, without accounting for the uncertainty in topology or branch lengths. For example, another paper in this issue (Wright, Brandley, and Hillis this issue) has estimated that the tree used to investigate the evolution of reproductive mode in reptiles (Pyron and Burbrink, 2013) has a mean bootstrap support of 86%, and that there are more than 1000 nodes in the tree with less than 50% bootstrap support. This degree of uncertainty could, in principle, lead to large overestimates of the number of evolutionary transitions between different modes of reproduction.

We re-iterate calls for ASR studies to account for uncertainty in tree topologies and branch lengths when making inferences about the evolutionary history of traits (Huelsenbeck and Bollback, 2001). If possible, it is preferable to use a joint-inference approach, such as the hierarchical Bayesian method (Huelsenbeck and Bollback, 2001) which allows uncertainty in the tree topology and branch lengths to be integrated out of inferences about ancestral states. Where this is not feasible, inferences should at the least be repeated on a large collection of trees that represent the uncertainty in the topology and branch lengths, to ensure as far as possible that point-estimates based on single tree are not misleading.

Our results highlight the more general issue that it can be difficult to make strong inferences from ASR studies without additional biological information. Models of trait evolution are necessarily oversimplified, and minor differences between models can lead to relatively major differences in conclusions. The bias we describe merely exacerbates these issues. While additional data and more sophisticated models of evolution might help to solve these problems, there is always likely to be plenty of room for debate. What is perhaps more useful are detailed discussions of the underlying biology of the trait(s) of interest. On that note, the detailed case studies of putative transitions from viviparity to oviparity presented by Pyron and Burbrink (2014) and Pyron et al. (this issue) are a welcome addition to the debate.

## ACKNOWLEDGMENTS

SD was supported by a Francisco José de Caldas Scholarship from the Colombian government and by a Sydney World Scholars Award from the University of Sydney. RL was supported by the Australian Research Council.

## APPENDIX

To set up the problem, imagine that we are interested in reconstructing the number of transitions between two states in a clade of $N$ species. We start with a true tree of $N$ species that is balanced and rooted, and that is constructed such that all of the species in the left-hand clade (clade 1) share one value of the trait, and all of the species in the right-hand clade (clade 2) share the other value of the trait. Figure 1A provides an example for 8 species.

We start by calculating the number of trees that will return the correct number of transitions, i.e. one. To estimate one transition, we require only that clade 1 and/or clade 2 is monophyletic. We start by counting the number of trees for which both clades are monophyletic. In this case, the number of possible trees is the square of the number of possible ways of rearranging a tree of $N/2$ species, since each arrangement of the $N/2$ species in clade 1 can be paired with any possible arrangement of the species in clade 2. The number of possible trees for a clade of $N/2$ species is taken from Felsenstein (Felsenstein, 2004), giving:

$$[(N-3)!!]^2 \tag{1}$$

We now count the number of trees for which just one clade is monophyletic, but the other paraphyletic, e.g. if clade 1 is monophyletic but the root of the tree appears in clade 2. In this case, we must account for all possible ways that the root of the tree could be placed within the paraphyletic clade. Since the two clades are both of size $N/2$ in this example, there are $N$-$2$ internal branches in each clade, so $N$-$2$ possible ways to root the tree in each clade. For each of these possible rootings of the paraphyletic clade, we must account for all possible ways to arrange the taxa in the monophyletic clade (of which there are *(N-3)!!*), and finally we account for the fact that either clade could be paraphyletic, giving:

$$2(N-2)(N-3)!! \tag{2}$$

Thus, the total number of trees which will estimate the correct number of transitions is the sum of equations (1) and (2), which simplifies to:

$$(N-3)!![(N-3)!! + (2N-4)] \tag{3}$$

All other trees will overestimate the number of transitions, because they will intermingle the species from clade 1 and clade 2 (e.g. Fig. 1B). Thus, the proportion of all possible trees that would overestimate the true number of transitions is the difference between equation 3 and the number of all possible rooted trees for $N$ taxa, divided by the latter:

$$\frac{(2N-3)!! - (N-3)!![(N-3)!! + (2N-4)]}{(2N-3)!!} \tag{4}$$

## LITERATURE CITED

Beaulieu JM, O'Meara BC, Donoghue MJ. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Syst Biol 62:725–737.

Beaulieu JM, Smith SA, Leitch IJ. 2010. On the tempo of genome size evolution in angiosperms. J Bot 2010:1–8.

Bennett TH, Flowers TJ, Bromham L. 2013. Repeated evolution of salt-tolerance in grasses. Biol Lett 9:20130029.

Bollback JP 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7:88.

Bouckaert R, Heled J, Kühnert D, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10: e1003537.

Bromham L, Bennett TH. 2014. Salt tolerance evolves more frequently in C4 grass lineages. J Evol Biol 27:653–659.

Diaz-Uriarte R, Garland T. 1998. Effects of branch length errors on the performance of phylogenetically independent contrasts. Syst Biol 47:654–672.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLOS Biol 4:e88.

Edwards EJ, Smith SA. 2010. Phylogenetic analyses reveal the shady history of C4 grasses. Proc Natl Acad Sci USA 107:2532–2537.

Elliot MG, Crespi BJ. 2009. Phylogenetic evidence for early hemochorial placentation in eutheria. Placenta 30:949–967.

Felsenstein J 2004. Inferring Phylogenies. 2nd Editio. Sunderland: Sinauer Associates.

FitzJohn RG, Maddison WP, Otto SP. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst Biol 58:595–611.

Ho SY, Duchêne S, Duchêne D. 2015. Simulating and detecting autocorrelation of molecular evolutionary rates among lineages. Mol Ecol Resour 15:688–696.

Huelsenbeck JP, Bollback JP. 2001. Empirical and hierarchical Bayesian estimation of ancestral states. Syst Biol 50:351–366.

Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic mapping of morphological characters. Syst Biol 52:131–158.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global diversity of birds in space and time. Nature 491:444–448.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. Mamm protein metab 3:21–132.

Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. PLoS Comput Biol 5:e1000520.

Odom KJ, Hall ML, Riebel K, Omland KE, Langmore NE. 2014. Female song is widespread and ancestral in songbirds. Nat Commun 5: Article number:3379.

Plummer M, Best N, Cowles K, Vines K. 2006. CODA: Convergence diagnosis and output analysis for MCMC. R news 6:7–11.

Pyron RA, Burbrink FT. 2013. Phylogenetic estimates of speciation and extinction rates for testing ecological and evolutionary hypotheses. Trends Ecol Evol 28:729–736.

Pyron RA, Burbrink FT. 2014. Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. Ecol Lett 17: 13–21.

Pyron RA, Wiens JJ. 2011. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol Phylogenet Evol 61: 543–583.

Rabosky DL, Santini F, Eastman J, et al. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat Commun 4.

Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Meth Ecol Evol 3: 217–223.

Revell LJ. 2014. Ancestral character estimation under the threshold model from quantitative genetics. Evolution (N Y) 68:743–759.

Stadler T 2011. Simulating trees with a fixed number of extant species. Syst Biol 60:676–684.

Zanne AE, Tank DC, Cornwell WK, et al. 2014. Three keys to the radiation of angiosperms into freezing environments. Nature 506:89–92.