Including autapomorphies is important for tip-dating with clocklike data, but not with non-clock data (#13158)

First submission

Please read the **Important notes** below, and the **Review guidance** on the next page.
When ready **submit online**. The manuscript starts on page 3.

Important notes

Editor and deadline

Laura Wilson / 30 Sep 2016

Files	2 Figure file(s) 2 Table file(s) 1 Other file(s) Please visit the overview page to **download and review** the files not included in this review pdf.
Declarations	No notable declarations are present

Review guidelines

Please in full read before you begin

How to review

When ready **submit your review online**. The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

🗋 You can also annotate this **pdf** and upload it as part of your review

To finish, enter your editorial recommendation (accept, revise or reject) and submit.

BASIC REPORTING

- ✓ Clear, unambiguous, professional English language used throughout.
- ✓ Intro & background to show context. Literature well referenced & relevant.
- ✓ Structure conforms to **PeerJ standard**, discipline norm, or improved for clarity.
- ✓ Figures are relevant, high quality, well labelled & described.
- ✓ Raw data supplied (See **PeerJ policy**).

EXPERIMENTAL DESIGN

- ✓ Original primary research within **Scope of the journal**.
- ✓ Research question well defined, relevant & meaningful. It is stated how research fills an identified knowledge gap.
- ✓ Rigorous investigation performed to a high technical & ethical standard.
- ✓ Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- ℹ Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- ✓ Data is robust, statistically sound, & controlled.
- ✓ Conclusion well stated, linked to original research question & limited to supporting results.
- ✓ Speculation is welcome, but should be identified as such.

The above is the editorial criteria summary. To view in full visit **https://peerj.com/about/editorial-criteria/**

Including autapomorphies is important for tip-dating with clocklike data, but not with non-clock data

Nicholas J Matzke Corresp., [1] , **Randy Irmis** [2]

[1] Division of Ecology, Evolution, and Genetics, Research School of Biology, The Australian National University, Canberra, Australia
[2] Department of Geology & Geophysics, University of Utah, Salt Lake City, Utah, United States

Corresponding Author: Nicholas J Matzke
Email address: nickmatzke.ncse@gmail.com

Tip-dating, where fossils are included as dated terminal taxa in Bayesian dating inference, is an increasingly popular method. Data for these studies often come from morphological character matrices originally developed for non-dated~~, and usually parsimony,~~ analyses. In parsimony, only shared derived characters (synapomorphies) provide grouping information, so many character matrices have an ascertainment bias: they leave out autapomorphies (~~unique derived character states~~), which are considered uninformative. There has been no study of the affect of this ascertainment bias in tip-dating, but autapomorphies can be informative in model-based inference. We expected that excluding autapomorphies would shorten the morphological branchlengths of terminal branches, and thus bias downwards the time branchlengths inferred in tip-dating. We tested for this effect using a matrix for Carboniferous-Permian eureptiles where all autapomorphies had been deliberately coded. Surprisingly, date estimates are virtually unchanged when autapomorphies are excluded, although we find large changes in morphological rate estimates and small effects on topological and dating confidence. We hypothesized that the puzzling lack of effect on dating was caused by the non-clock nature of the eureptile data. We confirm this explanation by simulating strict clock and non-clock datasets, showing that autapomorphy exclusion biases dating only for the clocklike case. A theoretical solution to ascertainment bias is computing the ascertainment bias correction ($\mathrm{M}k_{\mathrm{parsinf}}$), but we explore this correction in detail, and show that it is computationally impractical for typical datasets with many character states and taxa. Therefore we recommend that palaeontologists collect autapomorphies whenever possible when assembling character matrices.

Title: Including autapomorphies is important for tip-dating with clocklike data, but not with non-clock data

Matzke, Nicholas J.[1*]; Irmis, Randall B.[2]

[1]Division of Ecology, Evolution, and Genetics, Research School of Biology, The Australian National University, ACT 2601 AUSTRALIA

[2]Natural History Museum of Utah and Department of Geology & Geophysics, University of Utah, 301 Wakara Way, Salt Lake City, UT 84108-1214, USA

[*]Corresponding author. Email: nick.matzke@anu.edu.au

Abstract

Tip-dating, where fossils are included as dated terminal taxa in Bayesian dating inference, is an increasingly popular method. Data for these studies often come from morphological character matrices originally developed for non-dated, and usually parsimony, analyses. In parsimony, only shared derived characters (synapomorphies) provide grouping information, so many character matrices have an ascertainment bias: they leave out autapomorphies (unique derived character states), which are considered uninformative. There has been no study of the affect of this ascertainment bias in tip-dating, but autapomorphies can be informative in model-based inference. We expected that excluding autapomorphies would shorten the morphological branchlengths of terminal branches, and thus bias downwards the time branchlengths inferred in tip-dating. We tested for this effect using a matrix for Carboniferous-Permian eureptiles where all autapomorphies had been deliberately coded. Surprisingly, date estimates are virtually unchanged when autapomorphies are excluded, although we find large changes in morphological rate estimates and small effects on topological and dating confidence. We hypothesized that the puzzling lack of effect on dating was caused by the non-clock nature of the eureptile data. We confirm this explanation by simulating strict clock and non-clock datasets, showing that autapomorphy exclusion biases dating only for the clocklike case. A theoretical solution to ascertainment bias is computing the ascertainment bias correction ($\mathrm{M}k_{\mathrm{parsinf}}$), but we explore this correction in detail, and show that it is computationally impractical for typical datasets with many character states and taxa. Therefore we recommend that palaeontologists collect autapomorphies whenever possible when assembling character matrices.

Keywords

tip-dating, total-evidence dating, autapomorphies, parsimony, Eureptilia, BEASTmasteR

Introduction

~~In parsimony phylogenetic analyses~~, the only data informative for reconstructing the tree topology are those with grouping information: potentially shared, derived character states (synapomorphies; (Hennig et al. 1999)). An autapomorphy – a state unique to one Operational Taxonomic Unit (OTU; (Mishler 2005)) -- contributes one step to any possible topology. Therefore, autapomorphies are routinely excluded from further analysis in ~~cladistics~~ programs (e.g., the TNT *xinact* and *info* commands (Goloboff et al. 2008); the PAUP* *exclude* command (Swofford 2003)), and autapomorphic characters are often not even collected during assembly of a character-taxon matrix.

In model-based inference, autapomorphies *can* be informative (Lewis 2001; Wright & Hillis 2014), because autapomorphies contribute information about the overall rate of change in the character matrix and site-specific rate heterogenetity. In addition, tip-dating analyses might be expected to be particularly sensitive to autapomorphies: all autapomorphies occur on terminal branches by definition, so their exclusion will shorten the morphological branchlengths of terminal branches (and thus presumably their time branchlengths), and perhaps increase estimated branch-wise rate variation.

An alternative to inclusion of autapomorphies is ascertainment-bias correction, where the likelihood of unobservable character patterns, L_{unobs}, is calculated, and the likelihood of the observed data is normalized by dividing by 1- L_{unobs} (Felsenstein 1992; Lewis 2001). The two common corrections are the Markov-*k* model with an ascertainment bias correction for the unobservability of invariant characters (M*k*-variable-only, or M*k*v; (Lewis 2001)), and Markov-*k* with an ascertainment bias correction for parsimony-uninformative characters, Mk_{parsinf} (Allman et al. 2010; Ronquist & Huelsenbeck 2003). These corrections are options in MrBayes and can be implemented in Beast2 XML, but several studies briefly mention that the scalability and correctness of Mk_{parsinf} computations may be problematic (dos Reis et al. 2016; Koch & Holder 2012; Matzke 2016).

The effect inclusion/exclusion of autapomorphies and ascertainment-bias correction has not been studied in a tip-dating context. Datasets appropriate for doing so are rare because they need to systematically collect all autapomorphies, as well as dates for the OTUs. Müller and Reisz (2006) ~~(Müller & Reisz 2006)~~ constructed an all-fossil, morphological matrix of early eureptiles and tested the effect of inclusion/exclusion of autapomorphies in undated Bayesian inference, and recommended including autapomorphies. Lee and Palci (2015) discussed the importance of autapomorphies for tip-dating, but did not ~~conduct a test~~. We obtained dates for Müller and Reisz's taxa, and use the dataset to test the effects of autapomorphy inclusion. Surprisingly, no effect on dates was found. This might be due to the non-clocklike nature of the dataset, an explanation we confirm with a simulation study that shows autapomorphy exclusion biases terminal branchlength estimates when the data are highly clocklike, but not in a non-clock dataset. We also examine the Mk_{parsinf} correction and show that it scales poorly for characters with more than two states, limiting its usability.

Methods

Data. The morphological matrix was taken from Müller and Reisz (Mül~~ler & Reisz~~ 2006). The date ranges for OTUs were derived from the literature, following best practices guidelines (2012). Correlation between time and morphological branchlengths in a TNT parsimony analysis was used as a rough assessment of clocklike behavior.

Tip-dating eureptiles. Tip-dating in Beast2 (Bouckaert et al. 2014; Drummond & Bouckaert 2015) with Birth-Death-Serial Sampling (BDSS) or SA-BDSS (Sampled Ancestors) tree models (Gavryushkina et al. 2015; Gavryushkina et al. 2014) requires a specialized XML input file. To set this up, we used BEASTmasteR (Alexandrou et al. 2013; Matzke 2015; Matzke & Wright 2016), a set of R functions that convert NEXUS character matrices, an Excel file containing tip date ranges, and other priors and settings, into XML. Three different site models were used: M*k*, M*k*v, and Mk_{parsinf}. The summary Maximum Clade Credibility (MCC) trees were plotted with 95% highest posterior densities (HPDs) on inferred node (blue) and tip dates (red) using *BEASTmasteR* functions and custom R scripts. Mean node dates, node 95% HPD widths, posterior probabilities, and rates were compared between pairs of analyses (with/without autapomorphies) for nodes/bipartitions shared between analyses (n=14), with the Wilcoxon signed-rank test (WSRT) for paired samples. Due to the small number of tests, no multiple-test correction was used.

Simulation. To test whether clocklike behavior is needed to observe effects of autapomorphy exclusion on date estimates, a BDSS tree similar in size to the empirical dataset (30 species) was simulated using *TreeSim* (Stadler 2015). A “strict clock” dataset of 1000 binary characters was simulated on this tree under the M*k* model with a rate low enough that a substantial proportion of the characters were invariant or autapomorphic. A “non-clock” dataset was produced by reshuffling the time-branchlengths of the simulated tree, and then simulating another 1000 characters at the same rate. Datasets were filtered to produce variable-only and parsimony-informative-only datasets, effectively imposing ascertainment bias. Beast2 runs were conducted on both simulated datasets under M*k*, M*k*v, and Mk_{parsinf} using the same setup as for the empirical analysis. All scripts, Beast2 inputs and outputs, and further details of the analyses are available in Supplemental Material (SM).

*Scalability of the M*k$_{parsinf}$ *correction*. Although listed as an option in MrBayes for ~~a~~ over a decade, surprisingly, Mk_{parsinf} has not been formally described anywhere in the literature, leading to widespread lack of knowledge of how it works and whether or not it is computationally feasible on typical datasets. Nor has there been any formal treatment of its computational scalability. The key issue is the number of unobservable character patterns for a character with a particular number of states, as the likelihood of each unobservable pattern must be calculated. While this is feasible for a binary character (which appears to be the assumption made by MrBayes), for a dataset with many taxa and multistate characters, the number of unobservable site patterns rapidly climbs into the millions. Appendix 1 contains a derivation of the number of likelihood calculations required by Mk_{parsinf}, and discussion of computational scalability.

Results

Tip-dating eureptiles. Fourteen high-posterior bipartitions were inferred in all analyses. MCC trees for two runs are illustrated in Figure 1; for all runs, see Supplemental Figure 1. Summary statistics of key parameters are shown in Table 1. Linear regression of tip age against the root-to-tip distance in a parsimony analysis (the number of morphological steps on all branches leading to a tip) indicated that time and parsimony branchlengths were not correlated. This is evidence that the morphological characters in the eureptile dataset are not evolving in a clocklike manner.

Inferred node dates. Estimates of the root age are almost identical between analyses with and without autapomorphies (Table 1). Comparing mean dates for nodes shared across the MCC trees yields no significant differences (WTST, two-sided, *n*=14 shared nodes), with *P*=0.359 for the M*k* inference, and *P*=0.280 for M*k*v inferences.

Dating uncertainty (HPD widths). Adding data should reduce uncertainty, especially with small morphological datasets. The null hypothesis, that the no-autapomorphies dataset does not have greater HPD widths, was rejected for the M*k* inferences (including vs. excluding autapomorphies, 9.20 vs. 9.94, *P*= 0.023, one-sided WSRT), and a suggestive result for the M*k*v inferences (9.37 vs. 9.66, *P*=0.105).

Posterior probabilities (PPs). PPs were higher for runs including autapomorphies under both the M*k* model (including vs. excluding autapomorphies, 0.902 vs. 0.756) and the M*k*v model (0.900

vs. 0.835). The null hypothesis, that the no-autapomorphies dataset does not have smaller PPs, was rejected at a significance level of 0.05 for both the M*k* inference (*P*=0.0095, one-sided WSRT) and M*k*v inference (*P*=0.0252).

Relaxed clock. The mean of the relaxed clock rate is dramatically affected by inclusion of autapomorphies, under both the M*k* model (with autapomorphies, rate mean= 0.0782 changes per site per million years, 95% HPD=[0.015, 0.159]; without: 0.788 [0.0305, 3.982]) and the M*k*v model (with: 0.0376 [0.0074, 0.0840]; without: 0.550 [0.0228, 2.655]) (tests in SM). The Mk_{parsinf} run of the no-autapomorphies dataset yielded an intermediate clock rate (0.235, 95% HPD=[0.0142, 0.664]).

Simulations. Figure 2 shows the simulation procedure and key comparisons. Similar tree topologies were inferred under all datasets, but estimated time-branchlengths differed. When the characters are clocklike and autapomorphies are included, inferred time-branchlengths are highly accurate (2a). However, when autapomorphies are excluded, inferred terminal branchlengths are biased downwards (and accuracy decreases overall). The effect in 2b can also be seen by comparing inference while including vs. excluding autapomorphies, when the characters are clocklike (2c), but this effect disappears for non-clock data (2d).

Feasibility of $Mk_{parsinf}$. Equations in Appendix 1 demonstrate that $Mk_{parsinf}$ can be feasible for 2-state characters, and for 3-state characters on small datasets (~10 times slower for our dataset), but rapidly becomes computationally impractical as the number of taxa or states increases. The number of unobservable site patterns for various combinations of numbers of taxa and character states are shown in Table 2.

Discussion

Although estimated mean rate parameters for the eureptile dataset dropped dramatically when autapomorphies were included (and somewhat less when ascertainment-bias correction was used instead), the downstream effects on confidence were small (Table 1; SM), and there was no detectable effect on date inference. This seems surprising, because the exclusion of autapomorphies must reduce the number of morphological changes on terminal branches. However, reflection on the interaction between non-clocklike data, and the flexibility of relaxed-clock Bayesian tip-dating methods, provides an explanation. If the character data are non-clocklike, then the method will estimate a high rate of branchwise rate variation, indicating lack of correlation between time elapsed and morphological branchlength. In this situation, most of the dating information for the analysis comes from the serial-sampling of fossil tips rather than morphological branchlengths. If morphological branchlength is not correlated with time, this remains true whether or not autapomorphies are included, and adding autapomorphies is not likely to change the dating inference.

Our simulation results (Figure 2) confirm this explanation. The analysis of the empirical eureptile dataset is likely similar to the situation shown in Figure 2d: inferred time branchlengths are roughly the same whether or not autapomorphies are included. However, on a clocklike dataset, exclusion of autapomorphies clearly has an effect (Figure 2b). This suggests that the importance of including autapomorphies in tip-dating analyses depends on whether or not the characters have clocklike behavior. Unfortunately, assessing clocklike behavior will be more difficult when autapomorphies have been ignored or gathered only inconsistently (as is common).

An alternative to coding autapomorphies is the Mk_{parsinf} model. However, Appendix 1 shows that it scales too poorly to be generally useful for characters with large number of states (Table 2; SM). All versions of MrBayes back to at least 3.1.2 allow a "coding=informative" ascertainment bias correction to be specified, but the increase in computation time for a run with a single discrete character is very similar whether the character has 2, 3, 4, or 5 states (tested on MrBayes versions 3.1.2 through 3.2.6, and the 3.2.7 development version downloaded on September 7, 2016; data not shown). This suggests that Mk_{parsinf} may be implemented assuming only binary characters, and may be formally incorrect for multistate characters (as briefly noted by (dos Reis et al. 2016; Matzke 2016)), despite many usages in the literature. However, as most morphological datasets are dominated by binary characters, this issue may have limited impact on inference, and requires further study.

Conclusion

Our study indicates that the common practice of repurposing character matrices devised for parsimony and undated Bayesian analyses may not be sufficient in the world of Bayesian tip-dating. For higher quality datasets (many characters, clocklike behavior), the bias in dating introduced by ignoring autapomorphies may become significant. Additionally, ascertainment bias corrections are at present computationally impractical for many datasets with multistate characters. Finally, autapomorphies have additional utility for improving estimates of rates and rate variation, for species identification, for measuring disparity, and because autapomorphies may become synapomorphies when new taxa are described. Therefore, we recommend that autapomorphies be coded and used whenever possible.

Appendix 1: Derivation of the Mk_{parsinf} ascertainment bias correction, and problems with scalability

One potential alternative to our recommendation to code autapomorphies could be to employ the M*k*-Parsimony-Informative model (Mk_{parsinf}), that is, the Markov-*k* model with an ascertainment bias correction for not just the unobservability of invariant characters, but also for the unobservability of autapomorphic characters (in a dataset that excludes autapomorphic characters).

A significant question is whether or not the Mk_{parsinf} model can actually be employed on typical datasets. *MrBayes*, since at least version 3.1.2, does allow the Mk_{parsinf} ascertainment bias correction as an option (“lset coding=informative”; (Ronquist et al. 2011), p. 146, or http://mrbayes.sourceforge.net/wiki/index.php/Evolutionary_Models_Implemented_in_MrBayes_3#Standard_Discrete_.28Morphology.29_Model), but does not contain an extensive description of how it works, and the model does not seem to have been formally described in any publication. Allman et al. (~~Allman et al.~~ 2010) analyse model identifiability in the Mk_{parsinf}

context, but only cite Nylander et al. (Nylander et al. 2004) for the model; Nylander et al. in turn cite Ronquist and Huelsenbeck, “in prep.,” which appears to be an uncorrected reference to their cited (Ronquist & Huelsenbeck 2003), the *MrBayes 3* publication.

There is therefore a need to explore Mk_{parsinf} in detail. Ascertainment bias correction works by enumerating site patterns that are unobservable, calculating their likelihood under the model, and then re-normalizing the observed data likelihood at each site by dividing by 1-*L*, where *L* is the likelihood of the unobservable site pattern(s). In Felsenstein (~~Felsenstein~~ 1992), the unobservable pattern was “no restriction site observed,” i.e., a column of all 0s. In the M*kv* model, the unobservable patterns include a column of all 0s, a column of all 1s, etc., up to the number of character states, *n*, in a particular character.

However, the situation becomes much more complicated for the Mk_{parsinf} ascertainment bias correction (Supplemental Table 1). The difficulty (mentioned briefly in ~~(~~dos Reis et al. 2016; Matzke 2016)) is that the number of unobservable site patterns scales very poorly with number of character states and number of taxa. For example, for a 100 taxon data matrix and a 3-state character, the following is an unobservable site pattern: a column that consists of all 0s, a single 1 for taxon 99, and a single 2 for taxon 100. But any other variant of this pattern is also unobservable: all 0s, except state 1 at position 1, and state 2 at position 2, etc. Additional unobservable patterns include all 1s except for two taxa and all 2s except for two taxa. Also unobservable are all patterns that are invariant (all 0s, all 1s, all 2s), and all patterns that are invariant except for one taxon.

Formally speaking, if n is the number of taxa, and k is the number of states in a character, then there are k^n possible patterns for that character. We can calculate the number of site patterns that are unobservable under Mk_{parsinf} by first selecting the number of character states, i, found in a particular unobservable site pattern. For example, in a character assumed to have k=5 states, the number of states found in a particular pattern could be i=1 (i.e., an invariant site), i=2, …, i=5. For each i, there are

$$\binom{k}{i} \tag{1}$$

ways to pick which of the k character states will be found in found in the site pattern. Conditional on i character states in a particular unobservable site pattern, one of them will be the "dominant" state (held by all taxa, except for the autapomorphic taxa), and $(i-1)$ of the character states will be autapomorphies. There are

$$\binom{i}{1} \tag{2}$$

ways to choose which character state is dominant. Conditional on the dominant character state, there are

$$\binom{n}{i-1} \tag{3}$$

ways to choose which taxa will be autapomorphic. Conditional on which taxa are autapomorphic, there are ${}_{i-1}P_{i-1}$ permutations of ways to assign the $(i-1)$ character states to the $(i-1)$ autapomorphic taxa. This is calculated

$$\frac{(i-1)!}{((i-1)-(i-1))!} = \frac{(i-1)!}{0!} = (i-1)! \qquad (4)$$

Taking the product of equations *1-4* and summing over all *i* yields

$$\sum_{i=1}^{k} \binom{k}{i} \binom{i}{1} \binom{n}{i-1} (i-1)!$$

$$\sum_{i=1}^{k} \binom{k}{i} i! \binom{n}{i-1} \qquad (5)$$

autapomorphic patterns that are unobservable under the Mk_{parsinf} ascertainment bias correction, for a character with *k* states. This equation is implemented in the R function *num_unobservable_patterns_ParsInf*, included in the Supplemental Material (and available online via GitHub Gist, at: https://gist.github.com/nmatzke/8f80723b6e1fc80ed5ac).

Calculating the number of unobservable patterns for a range of numbers of taxa and states (Supplemental Table 1) shows that, for a 100-taxon morphological matrix, the presence of characters with 3 states in the matrix will necessitate calculating the likelihood for 30,303 additional site patterns. This is computationally imaginable, although it will substantially slow the MCMC search for a morphological dataset, which usually has only a few hundred characters. The presence of a 4-state character requires 4,000,804 unobservable patterns. For a 6-state character there are over 57 billion.

In *Beast2*, unobservable site patterns have to be physically listed in the XML input; even with a script to write out the patterns, users can certainly imagine the difficulty of saving and manipulating XML files containing millions of unobservable patterns. Inspection of the *MrBayes* code seems to indicate that the Mk_{parsinf} correction assumes binary characters only (which is computationally feasible; Supplemental Table 1); but this leaves open the question of what calculation, is being done on characters with more than two states.

This is problematic, as many researchers (e.g. ~~(~~Dembo et al. 2016~~)~~) are probably under the impression that Mk_{parsinf} ascertainment bias correction works for any number of character states. It is possible that this issue is of little significance. After all, most morphological characters are binary. Also, as the number of taxa and character states increases, the fraction of the total number of possible patterns (k^n) that are unobservable (equation *5*) decreases precipitously (Supplemental Tables 2, 3). Thus, perhaps the likelihood of unobservable sites dwindles to irrelevance. This seems to be the observation made in the MrBayes manual ((Ronquist et al. 2011), pp. 146-147), where the authors state they observed that as the total tree length (sum of branchlengths in terms of number of expected changes per site) increases e.g. above 20-30 taxa, the ascertainment bias correction becomes negligible. However, this may depend greatly on the “true” rates – if they are low, and thus invariant and autapomorphic patterns are some of the most probable patterns, then the likelihood correction from unobservable patterns could be large. As this paper showed, in the case of the 25-taxon eureptile dataset, just switching from M*k* to M*k*v models dropped the mean clock rate estimate by about 1/3 in both the autapomorphies-included and autapomorphies-excluded dataset. This would affect the morphological branchlengths (number of expected changes per site) in a similar way. As Mk_{parsinf} includes the M*k*v correction, this suggests Mk_{parsinf} would have the same or greater effect.

Two comments we received from readers of a draft of this manuscript deserve attention. First, Mike Lee (personal communication) pointed out that the equations above would be somewhat different if the researchers coding characters excluded not only parsimony-uninformative characters, but also characters that were "partially uninformative." An example would be the character pattern 00112 – character states 0 and 1 are potential synapomorphies, but character state 2 is an autapomorphy. Above, we have focused on the "literal" interpretation of "parsimony-informative," which we think is the understanding commonly used in the literature and in programs. We suggest that an ascertainment bias correction that assumes the unobservability of invariant, parsimony-uninformative, and partially-parsimony-informative characters should have a new name, perhaps simply "partial-parsinf."

Second, Mark Holder (personal communication) pointed out that the scalability problem is less detrimental, although still daunting, if it is realized that some patterns will have the same likelihood under the M*k* model (because it is a symmetric-equal-rates model). For example, the patterns 00112, 00221, 11002, 11220, 22110, and 22001, will all have the same likelihood. Therefore the log-likelihood can be calculated for one of these patterns, and multiplied by the number of patterns in that category. This amounts to removing $i!$ from equation (5), and using it as a weight to multiply by the log-likelihood of a pattern. Beast2 does have a "weight" option for its Alignment class, but we have not tested it in combination with the ascertained/excludefrom/excludeto options in the XML.

Equation (5) applies to unordered characters, where any autapomorphies will be parsimony-uninformative. If it is instead assumed that the characters are ordered, then any pattern with more than two states will be parsimony-informative. For example, the pattern 011112 would be parsimony-uninformative for an unordered character, but parsimony-informative for an ordered character, because bipartitions grouping states (0,1) and (1,2) would be favoured over trees grouping (0,2). Thus, the number of unobservable patterns (assuming the researchers doing the character scoring had this in mind when building their matrix) is much reduced, since only patterns with 1 or 2 character states are unobservable. The equation is:

$$\sum_{i=1}^{2}\binom{k}{i}\binom{i}{1}\binom{n}{i-1}(i-1)! \qquad (6)$$

The unobservable pattern counts for an ordered character are shown in Supplemental Table 4, and fractions in Supplemental Table 5.

Resolution of the discussion about when and where Mk_{parsinf} is functional, useful, or unnecessary may be difficult, as it depends in part on gnarly philosophical questions about what the “complete” morphology matrix would look like (how many invariant morphological characters are “truly” observable in any particular clade?). This is closely tied to another difficult question: what is the "true" morphological rate for “all” of the morphology? We can briefly suggest that probably such questions are almost unanswerable in the abstract, and that any meaningful statements about rates and completeness must be made with reference to some method of character collection. It certainly appears that these problems should be studied more carefully than can be done here. Unless these issues are resolved, however, it may be that including all codeable autapomorphies, and using the M*k*v ascertainment bias correction, is the best option.

Data accessibility

All data, code, and Beast2 inputs and outputs are available in this article’s Dryad data repository: http://datadryad.org/review?doi=doi:10.5061/dryad.8q4c8

Competing interests

We have no competing interests.

Authors’ Contributions

RBI conceived of the idea, obtained the dataset, and assembled the tip dates and documentation from the literature. NJM wrote *BEASTmasteR*, conducted the computational analyses and drafted the manuscript. Both authors planned the study, and edited and corrected the manuscript.

Acknowledgements

The authors also thank: April Wright, David Bapst, Graeme Lloyd, Daniel Ksepka, James Parham, Nathan Smith, Alan Turner, and Mike Lee for helpful discussions.

References

Alexandrou MA, Swartz BA, Matzke NJ, and Oakley TH. 2013. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Molecular Phylogenetics and Evolution* 69:514-523. 10.1016/j.ympev.2013.07.026

Allman ES, Holder MT, and Rhodes JA. 2010. Estimating trees from filtered data: Identifiability of models for morphological phylogenetics. *Journal of Theoretical Biology* 263:108-119. 10.1016/j.jtbi.2009.12.001

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, and Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10:e1003537. 10.1371/journal.pcbi.1003537

Dembo M, Radovčić D, Garvin HM, Laird MF, Schroeder L, Scott JE, Brophy J, Ackermann RR, Musiba CM, de Ruiter DJ, Mooers AØ, and Collard M. 2016. The evolutionary relationships and age of Homo naledi: An assessment using dated Bayesian phylogenetic methods. *Journal of Human Evolution* 97:17-26. http://dx.doi.org/10.1016/j.jhevol.2016.04.008

dos Reis M, Donoghue PCJ, and Yang Z. 2016. Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews: Genetics* 17:71-80. 10.1038/nrg.2015.8

Drummond AJ, and Bouckaert RR. 2015. *Bayesian Evolutionary Analysis with BEAST*: Cambridge University Press.

Felsenstein J. 1992. Phylogenies from restriction sites: a maximum-likelihood approach. *Evolution* 46:159-173. 10.2307/2409811

Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, and Drummond AJ. 2015. Bayesian total evidence dating reveals the recent crown radiation of penguins. *arXiv*. arXiv:1506.04797

Gavryushkina A, Welch D, Stadler T, and Drummond AJ. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput Biol* 10:e1003919. 10.1371/journal.pcbi.1003919

Goloboff PA, Farris JS, and Nixon KC. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24:774-786. 10.1111/j.1096-0031.2008.00217.x

Hennig W, Davis DD, and Zangerl R. 1999. *Phylogenetic Systematics*: University of Illinois Press.

Koch JM, and Holder MT. 2012. An Algorithm for Calculating the Probability of Classes of Data Patterns on a Genealogy. *PLOS Currents Tree of Life*. 10.1371/4fd1286980c08

Lee Michael SY, and Palci A. 2015. Morphological phylogenetics in the genomic age. *Current Biology* 25:R922-R929. 10.1016/j.cub.2015.07.009

Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913-925. 10.1080/106351501753462876

Matzke NJ. 2015 BEASTmasteR: automated conversion of NEXUS data to BEAST2 XML format, for fossil tip-dating and other uses. PhyloWiki. http://phylo.wikidot.com/beastmaster

Matzke NJ. 2016. The evolution of antievolution policies after *Kitzmiller versus Dover*. *Science* 351:28-30. 10.1126/science.aad4057

Matzke NJ, and Wright AM. 2016. Inferring node dates from tip dates in fossil Canidae: the importance of tree priors. *Biology Letters* 12:20160328. 10.1098/rsbl.2016.0328

Mishler BD. 2005. The logic of the data matrix in phylogenetic analysis. In: Albert VA, ed. *Parsimony, phylogeny, and genomics*. Oxford: Oxford University Press, 57-70.

Müller J, and Reisz RR. 2006. The phylogeny of early eureptiles: comparing parsimony and Bayesian approaches in the investigation of a basal fossil clade. *Systematic Biology* 55:503-511. 10.1080/10635150600755396

Nylander JAA, Ronquist F, Huelsenbeck JP, and Nieves-Aldrey J. 2004. Bayesian phylogenetic analysis of combined data. *Systematic Biology* 53:47-67. 10.1080/10635150490264699

Parham JF, Donoghue PCJ, Bell CJ, Calway TD, Head JJ, Holroyd PA, Inoue JG, Irmis RB, Joyce WG, Ksepka DT, Patané JSL, Smith ND, Tarver JE, van Tuinen M, Yang Z, Angielczyk KD, Greenwood JM, Hipsley CA, Jacobs L, Makovicky PJ, Müller J, Smith KT, Theodor JM, Warnock RCM, and Benton MJ. 2012. Best practices for justifying fossil calibrations. *Systematic Biology* 61:346-359. 10.1093/sysbio/syr107

Ronquist F, Huelsenbeck J, and Teslenko M. 2011 Draft MrBayes version 3.2 manual: tutorials and model summaries. SourceForge. http://mrbayes.sourceforge.net/mb3.2_manual.pdf

Ronquist F, and Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574. 10.1093/bioinformatics/btg180

Stadler T. 2015 TreeSim: Simulating Phylogenetic Trees. https://CRAN.R-project.org/package=TreeSim

Swofford DL. 2003 PAUP*. Phylogenetic Analysis Using Parsimony, version 4. http://paup.csit.fsu.edu/

Wright AM, and Hillis DM. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS ONE* 9:e109210. 10.1371/journal.pone.0109210

Figure 1(on next page)

Comparison of the tip-dated phylogenies of early eureptiles inferred when excluding or including autapomorphies.

Figure 1. Comparison of the tip-dated phylogenies of early eureptiles inferred when excluding (a) or including (b) autapomorphies, under M*k*v ascertainment bias correction. Numbers are posterior probabilities. Bars represent the 95% HPD.

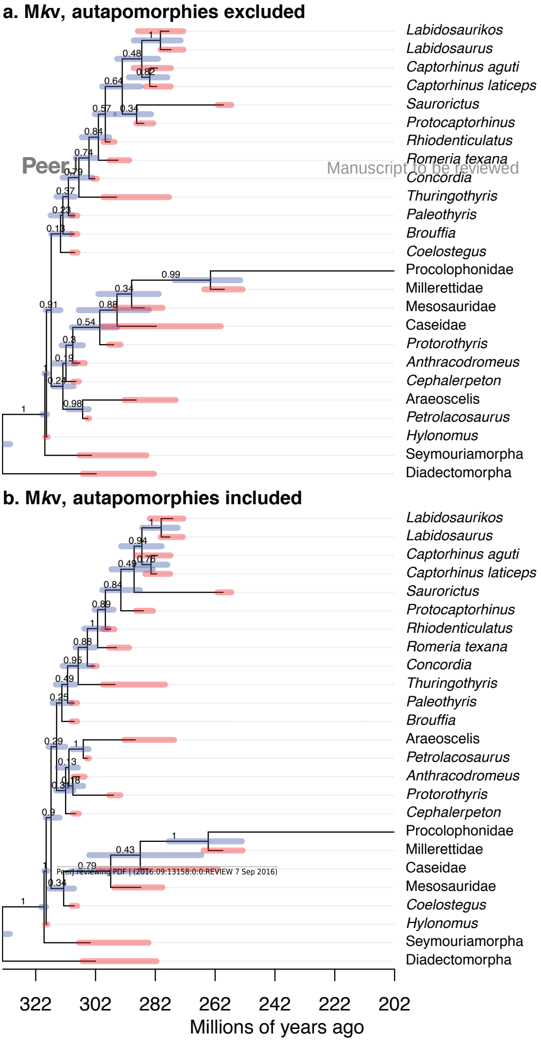
a. Mkv, autapomorphies excluded
Labidosaurikos
Labidosaurus
Captorhinus aguti
Captorhinus laticeps
Sauroctus
Protocaptorhinus
Rhiodenticulatus
Romeria texana
Concordia
Thuringothyris
Paleothyris
Brouffia
Coelostegus
Procolophonidae
Millerettidae
Mesosauridae
Caseidae
Protorothyris
Anthracodromeus
Cephalerpeton
Araeoscelis
Petrolacosaurus
Hylonomus
Seymouriamorpha
Diadectomorpha
1
0.48
0.82
0.64
0.57
0.34
0.84
0.74
0.79
0.37
0.23
0.13
0.99
0.34
0.91
0.88
0.54
0.3
0.19
1
0.24
0.98
1
b. Mkv, autapomorphies included
Labidosaurikos
Labidosaurus
Captorhinus aguti
Captorhinus laticeps
Sauroctus
Protocaptorhinus
Rhiodenticulatus
Romeria texana
Concordia
Thuringothyris
Paleothyris
Brouffia
Araeoscelis
Petrolacosaurus
Anthracodromeus
Protorothyris
Cephalerpeton
Procolophonidae
Millerettidae
Caseidae
Mesosauridae
Coelostegus
Hylonomus
Seymouriamorpha
Diadectomorpha
1
0.94
0.76
0.49
0.84
0.89
1
0.88
0.95
0.49
0.25
0.29
1
0.13
0.18
0.31
0.9
1
0.43
0.79
1
0.34
1
322
302
282
262
242
222
202
Millions of years ago

Figure 2(on next page)

Simulation procedure and results.

Figure 2. Simulation procedure (top) and results (bottom, a-d). The lack of an effect of excluding autapomorphies on dating in the empirical eureptile result is similar to the result on non-clock data shown in 2d.

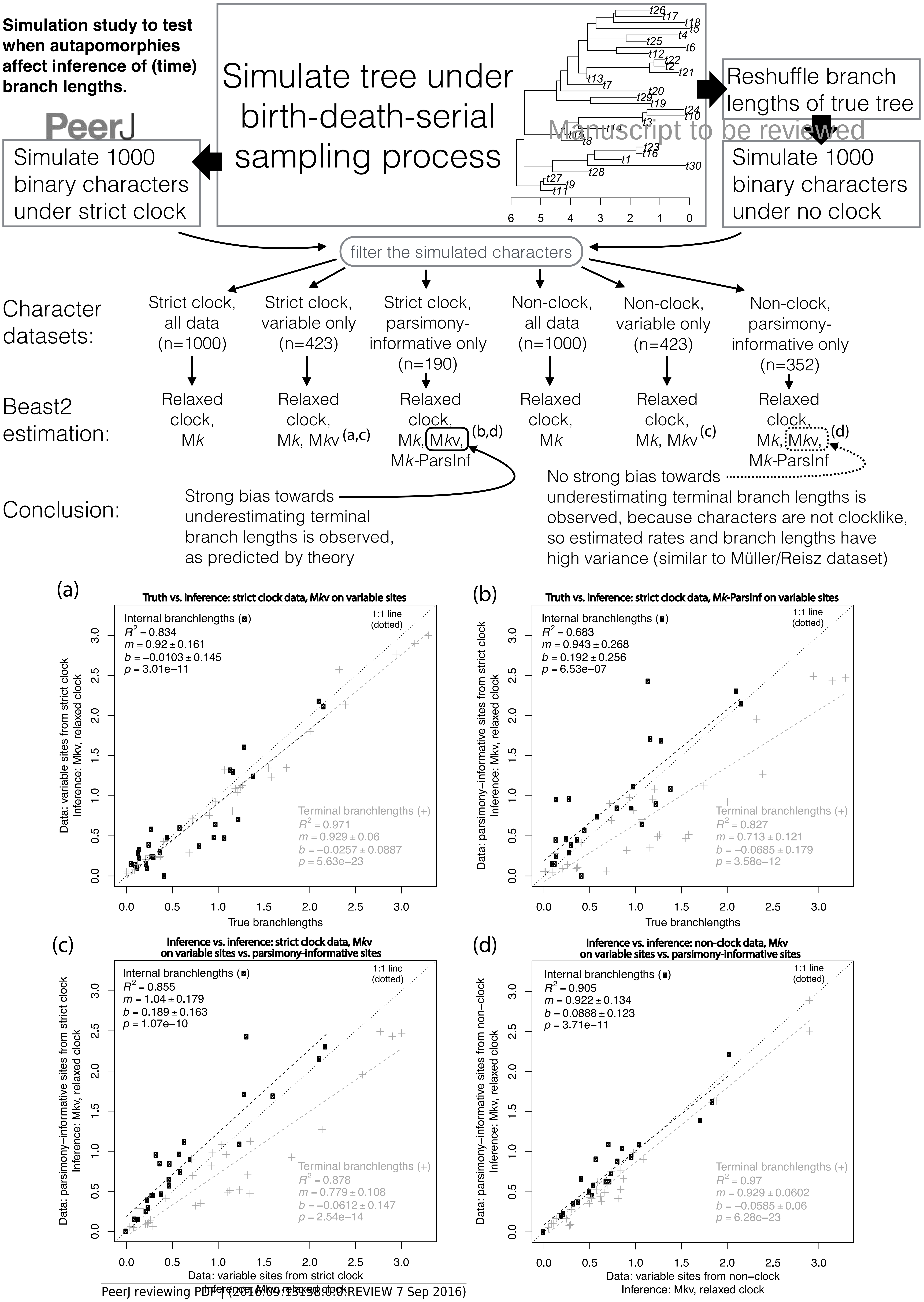
Simulation study to test when autapomorphies affect inference of (time) branch lengths.
Simulate tree under birth-death-serial sampling process
t26 t17 t18 t5 t4 t25 t6 t12 t22 t2 t21 t13 t7 t20 t29 t19 t24 t10 t3 t14 t15 t8 t23 t16 t1 t30 t28 t27 t9 t11
6 5 4 3 2 1 0
Simulate 1000 binary characters under strict clock
Reshuffle branch lengths of true tree
Simulate 1000 binary characters under no clock
filter the simulated characters
Character datasets:
Strict clock, all data (n=1000)
Strict clock, variable only (n=423)
Strict clock, parsimony-informative only (n=190)
Non-clock, all data (n=1000)
Non-clock, variable only (n=423)
Non-clock, parsimony-informative only (n=352)
Beast2 estimation:
Relaxed clock, Mk
Relaxed clock, Mk, Mkv (a,c)
Relaxed clock, Mk, Mkv, Mk-ParsInf (b,d)
Relaxed clock, Mk
Relaxed clock, Mk, Mkv (c)
Relaxed clock, Mk, Mkv, Mk-ParsInf (d)
Conclusion:
Strong bias towards underestimating terminal branch lengths is observed, as predicted by theory
No strong bias towards underestimating terminal branch lengths is observed, because characters are not clocklike, so estimated rates and branch lengths have high variance (similar to Müller/Reisz dataset)
(a)
Truth vs. inference: strict clock data, Mkv on variable sites
Internal branchlengths (■)
$R^2 = 0.834$
$m = 0.92 \pm 0.161$
$b = -0.0103 \pm 0.145$
$p = 3.01e-11$
1:1 line (dotted)
Terminal branchlengths (+)
$R^2 = 0.971$
$m = 0.929 \pm 0.06$
$b = -0.0257 \pm 0.0887$
$p = 5.63e-23$
Data: variable sites from strict clock
Inference: Mkv, relaxed clock
True branchlengths
0.0 0.5 1.0 1.5 2.0 2.5 3.0
(b)
Truth vs. inference: strict clock data, Mk-ParsInf on variable sites
Internal branchlengths (■)
$R^2 = 0.683$
$m = 0.943 \pm 0.268$
$b = 0.192 \pm 0.256$
$p = 6.53e-07$
1:1 line (dotted)
Terminal branchlengths (+)
$R^2 = 0.827$
$m = 0.713 \pm 0.121$
$b = -0.0685 \pm 0.179$
$p = 3.58e-12$
Data: parsimony-informative sites from strict clock
Inference: Mkv, relaxed clock
True branchlengths
0.0 0.5 1.0 1.5 2.0 2.5 3.0
(c)
Inference vs. inference: strict clock data, Mkv on variable sites vs. parsimony-informative sites
Internal branchlengths (■)
$R^2 = 0.855$
$m = 1.04 \pm 0.179$
$b = 0.189 \pm 0.163$
$p = 1.07e-10$
1:1 line (dotted)
Terminal branchlengths (+)
$R^2 = 0.878$
$m = 0.779 \pm 0.108$
$b = -0.0612 \pm 0.147$
$p = 2.54e-14$
Data: parsimony-informative sites from strict clock
Inference: Mkv, relaxed clock
Data: variable sites from strict clock
Inference: Mkv, relaxed clock
0.0 0.5 1.0 1.5 2.0 2.5 3.0
(d)
Inference vs. inference: non-clock data, Mkv on variable sites vs. parsimony-informative sites
Internal branchlengths (■)
$R^2 = 0.905$
$m = 0.922 \pm 0.134$
$b = 0.0888 \pm 0.123$
$p = 3.71e-11$
1:1 line (dotted)
Terminal branchlengths (+)
$R^2 = 0.97$
$m = 0.929 \pm 0.0602$
$b = -0.0585 \pm 0.06$
$p = 6.28e-23$
Data: parsimony-informative sites from non-clock
Inference: Mkv, relaxed clock
Data: variable sites from non-clock
Inference: Mkv, relaxed clock
0.0 0.5 1.0 1.5 2.0 2.5 3.0

Table 1(on next page)

Comparison of summary statistics from the five Beast2 runs.

Table 1. Comparison of summary statistics from the five Beast2 runs using "best-practices" tip dates.

Table 1. Comparison of summary statistics from the five Beast2 runs using "best-practices" tip dates.

Run #	**1**	**2**	**3**	**4**	**5**
Data	Including autapomorphies		Excluding autapomorphies		
Model	M*k*	M*k*v	M*k*	M*k*v	M*k*-parsinf
Ln Posterior	-1393.4	-1362.2	-1154.2	-1144.9	-1134.4
ESS	1801	1485	1801	1801	1801
Root age	332.6 [330.2, 335.3]	332.5 [330.0, 335.1]	332.6 [330.1, 335.1]	332.6 [330.1, 335.1]	332.6 [330.0, 335.1]
Birth	0.360 [0.0355, 1.316]	0.424 [0.0405, 1.708]	0.342 [0.0463, 1.221]	0.381 [0.0402, 1.377]	0.564 [0.0444, 2.841]
Death	0.336 [9.17e-5, 1.315]	0.3995 [1.13e-4, 1.723]	0.318 [4.97e-6, 1.220]	0.357 [2.57e-4, 1.391]	0.541 [6.37e-4, 2.843]
Sampling	0.0271 [7.90e-4, 0.0626]	0.0264 [0.00104, 0.0650]	0.0271 [8.85e-4, 0.063]	0.0261 [9.96e-4, 0.0634]	0.0256 [7.66e-4, 0.0643]
Clock rate mean	0.0782 [0.015, 0.159]	0.0376 [0.0074, 0.0840]	0.788 [0.0305, 3.982]	0.550 [0.0228, 2.655]	0.235 [0.0142, 0.664]
Clock rate SD	1.747 [1.201, 2.399]	1.712 [1.111, 2.309]	2.436 [1.572, 3.477]	2.341 [1.488, 3.379]	2.079 [1.318, 2.984]

Table 2(on next page)

Number of patterns unobservable under parsimony-informative ascertainment bias.

Table 2. Number of patterns that are unobservable under the Mk_{parsinf} ascertainment bias correction.

Table 2. Number of patterns that are unobservable in the $\mathrm{M}k_{\text{parsinf}}$ model.

		# states: 2	**3**	**4**	**5**	**6**
# of taxa	**4**	10	63	292	1045	3006
	5	12	93	544	2505	9276
	10	22	333	4084	42505	381546
	20	42	1263	32164	730005	15085086
	50	102	7653	500404	30062505	1698527706
	100	202	30303	4000804	490250005	57089105406
	200	402	120603	32001604	7921000005	1.87E+12
	500	1002	751503	500004004	3.11E+11	1.86E+14
	1000	2002	3003003	4000008004	4.99E+12	5.97E+15