Syst. Biol. 70(1):197–201, 2021
© The Author(s) 2020. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/syaa067
Advance Access publication August 26, 2020

Nothing Wrong with the Analysis of Clades in Comparative Evolutionary Studies: A Reply to Poe et al.

Joanna Baker, Andrew Meade, Mark Pagel, and Chris Venditti*

School of Biological Sciences, University of Reading, Reading, Berkshire, RG6 6BX, UK
**Correspondence to be sent to: School of Biological Sciences, University of Reading, Reading, Berkshire, RG6 6BX, UK; E-mail: c.d.venditti@reading.ac.uk.*

Received 26 April 2020; reviews returned 00 Month 2020; accepted 15 August 2020
Associate Editor: Luke Harmon

Abstract.—In a recent paper, Poe et al. assert that scientists should abandon clade-based approaches, particularly those using named taxonomic ranks. Poe et al. attempt to demonstrate that clade selection can have effects on the results of evolutionary analyses but unfortunately fall short of making any robust conclusions. Here, we demonstrate that the assertions made by Poe et al. have two important flaws: (i) an erroneous view of modern phylogenetic comparative methods; and (ii) a lack of statistical rigor in their analyses. We repeat Poe et al.'s analysis but using appropriate phylogenetic comparative approaches. We demonstrate that results remain consistent regardless of the clade definition. We go on to discuss the value of taxonomic groupings and how they can provide meaningful units of comparison in evolutionary study. Unlike the disheartening suggestion to abandon the use of clades, scientists can instead continue to use phylogenetic "*corrections*" that are already the standard for most comparative evolutionary analyses. [Comparative methods; evolution; phylogeny; taxonomy.]

On the Selection and Analysis of Clades: Poe et al. (2020)

Poe et al. (2020) critique the use of clade comparisons in biological analyses, arguing that owing to human subjectivity in taxonomic ranks, their use can introduce biases. They describe four alternative methods for defining clades and discuss problems with each—including subjectivity and phylogenetic nonindependence. Poe et al. (2020) then attempt to demonstrate that using alternative clade definitions can alter the results of evolutionary analyses.

Nonindependence among Clades

Over the last 30 years, biologists have come to understand the importance of accounting for nonindependence among taxa owing to shared ancestry in our evolutionary analyses (Felsenstein 1985; Pagel 1999). Poe et al. (2020) recognize that this argument extends to clades as well as species—suggesting approaches that study clades (of any definition) do not account for nonindependence attributable to shared ancestry. However, most modern phylogenetic comparative approaches employ methods that account for the nonindependence of clades. For example, phylogenetic generalized least squares methods for continuously varying characters (Martins and Hansen 1997; Pagel 1997, 1999) as well as phylogenetic generalized linear mixed models for various other data types (e.g., Hadfield and Nakagawa 2010) transform the underlying phylogeny into a variance–covariance (VCV) matrix that explicitly represents the evolutionary correlation among all lineages. Any clade, no matter how it is defined, is included within the VCV matrix that represents the entire phylogeny.

It is true that separately performing analyses on individual clades will suffer from biases associated with nonindependence (even where those analyses are phylogenetic), but only where explicit statistical comparisons between clades are made (e.g., using clade means). However, seeking evolutionary comparisons across clades by conducting comparative analyses on multiple monophyletic clades (e.g., Sánchez-Reyes et al. 2017; González-del-Pliego et al. 2019) is valid. There would, for example, be nothing statistically wrong with taking the data from an experiment, then analyzing the main effects in each block separately: there is a loss of power, and controls must be included for multiple testing, but there would be no inherent bias. In fact, making comparisons among subclades of a larger tree can prove invaluable, particularly where large timescales are involved and thus where evolutionary patterns and processes are unlikely to be constant. Revealing the nuances of evolution among clades can be made possible by clade comparisons.

Defining Clades

Poe et al. (2020) describe four main ways of defining clades for comparison within evolutionary analyses: (i) named taxonomic ranks such as families, orders, etc. (henceforth *taxonomic clades*); (ii) randomly selected subclades from the wider phylogeny (*random clades*); (iii) time-slicing the phylogeny to produce clades of equal or similar age (*time-sliced clades*); (iv) sister clades defined using a variety of approaches including equalizing trait/diversity similarity, ensuring similar sample size, and randomization (*sister clades*). After understanding that it is possible to account for nonindependence amongst clades of any definition (see above), Poe et al.'s criticism of taxonomic ranks appears to be human subjectivity. However, other methods are not immune to this issue, including the time-sliced clades, which seem to be preferred by Poe et al. For example, how is a time slice chosen? Are geological

time stages or major evolutionary events considered? Furthermore, most standard evolutionary analyses are implemented within a Brownian framework, which is scale invariant; using time-sliced clade definitions, statistically speaking, is no different to any other clade definition.

One proposed solution to reduce subjectivity in the use of named taxa is to use phylogenetically defined clades (Poe et al. 2020). Whilst the use of phylogeny to name taxa is not itself without controversy (e.g., Cantino and De Queiroz 2000; Forey 2002), modern taxonomy is constantly updated in the face of new phylogenetic data. For example, Hippopotamidae has since been moved from the artiodactyl suborder Suiformes on the basis of revised molecular and phylogenetic information and is now accepted to form a monophyletic clade with modern cetaceans: Whippomorpha (Waddell et al. 1999). Taxonomic groups are therefore often phylogenetic by definition and have also been shown to be stable and informative units for evolutionary analysis (Humphreys and Barraclough 2014; Barraclough and Humphreys 2015).

Effect of Clade Choice on Evolutionary Analyses

Poe et al. (2020) attempted to use the analysis and approach of Baker et al. (2015) to demonstrate differences in results of evolutionary analyses when time-sliced clades are used as opposed to taxonomic clades. The original analysis of Baker et al. (2015) detected the rate of body size evolution across the mammalian phylogenetic tree (Fritz et al. 2009) using the variable rates model (Venditti et al. 2011). This model produces a phylogenetic tree with branches measured by the rate of evolution—where branches are longest, body size changed more rapidly during the course of mammalian evolution. Across all mammals, the rate of body size evolution (measured as the sum of rate-scaled branches along a species' path) was associated with larger body size (Baker et al. 2015), that is, where the rate of body size change was fastest it tended to be towards a larger size. The relationship within clades of mammals corresponding to the phylogenetic groupings defined by Bininda-Emonds et al. (2007) was then compared by including a separate intercept and slope (interaction between clade and path length) for all clades in one single statistical model (Baker et al. 2015) that included a VCV matrix.

Here, we show that the conclusions of Poe et al. are unjustified owing to several flaws in their analysis including (i) a potentially biased data set, (ii) independent treatment of clades, and (iii) apparently nonphylogenetic analyses across species (see Methods for details). We conduct two new sets of phylogenetic generalized least squares regressions determining the relationship between body size and path-wise rate within (i) the taxonomic clades identified by Bininda-Emonds et al. (2007) and (ii) time-slice clades identified using a time slice of 86 MYA (Poe et al. 2020, Fig. 1). All analyses herein are conducted using the rates and data set ($N = 3321$) from Baker et al. (2015).

In both cases, the majority of groups studied show a significant ($P < 0.05$) relationship (Fig. 2, Table 1). The only minor difference in the analyses between clade definitions is within primates: neither strepsirrhine nor haplorrhine primates have a significant slope. Primates were one of the least significant slopes in the original analysis (Baker et al. 2015), and partitioning that variance into two subgroups may have further diminished signal. Additionally, bursts of body size increase occur on several key branches of the primate phylogeny including those leading to ancestral primates, strepsirrhines, and haplorhines (Baker et al. 2015)—such branches are explicitly not included in comparisons dividing primates into two groups (Poe et al. 2020, Fig. 2b). The body size increase is not universal, but instead, there is a directional bias along individual branches where rates are highest. This does not invalidate the interpretation of a trend towards increasing body size across primates but could point towards a causal explanation.

The above issues notwithstanding, Poe et al. (2020) test the effect of clade definition on the results of evolutionary models without providing statistical support for whether one definition is preferred over another. Here, we find that there is strong support (Raftery 1996) for using taxonomic groupings over time-slice groupings (Bayes Factor $= 69.78$, see Materials and Methods). Despite this result, we do not claim that clade comparisons based on named taxonomic groups are the only way to study biological evolution. In fact, the choice of clade definition for analyses seeking to detect generality of patterns across long evolutionary timescales may be mostly irrelevant. For example, the key message of Baker et al. (2015) was that there is a long-term evolutionary bias towards increasing size in the mammals. In our results here, we find that significantly more individual branches of the mammalian phylogeny increase in size regardless of how clades are defined ($P < 0.001$, see Materials and Methods): 66.8% of all lineages increase in body size when we study taxonomic clades compared with 64.2% for time-sliced clades (calculated using ancestral state inferences, see methods). There are also often other things to consider when choosing which clades to study, including sample size and whether there is sufficient within-clade variation for statistical analysis (phenotypic, ecological, or geological etc.). In any case, the conclusions of Baker et al. (2015) remain unchanged regardless of which way clades are defined (Fig. 2, binomial test results above): there is a general tendency for body size increase throughout mammalian evolutionary history.

Conclusions

The perceived inherent nonindependence bias of most clade-based approaches presented by Poe et al. (2020) is based on misunderstandings about modern phylogenetic comparative methods. There is no general

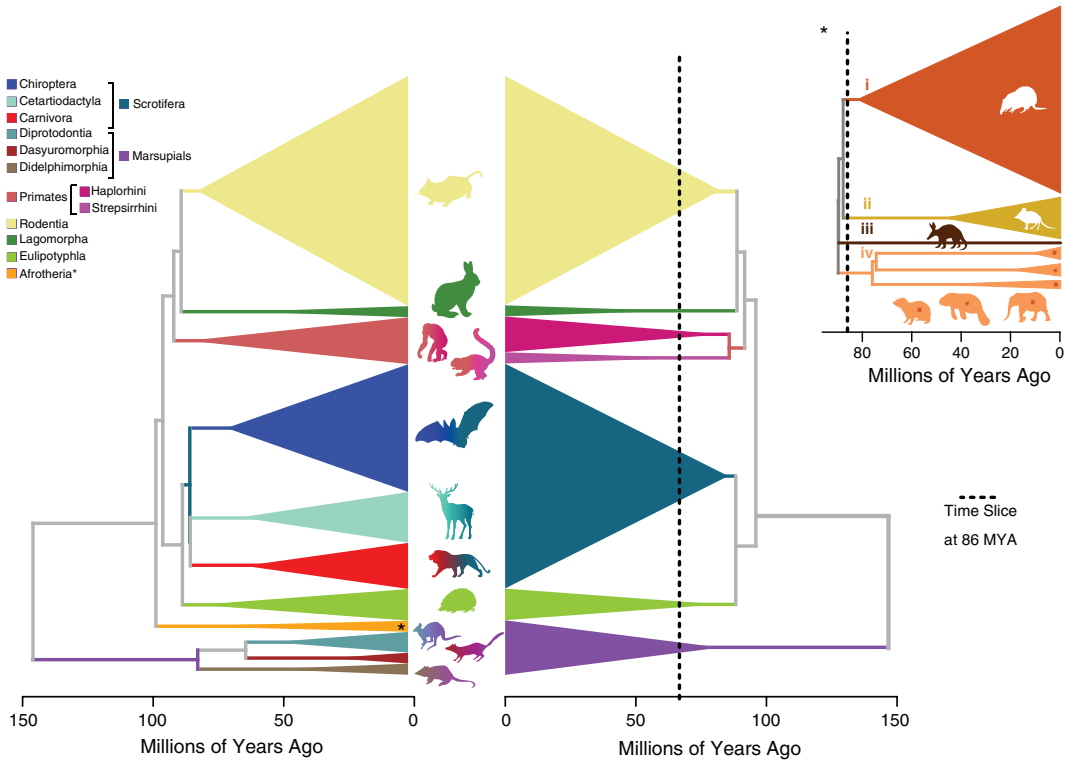

FIGURE 1. Comparison between taxonomic clades and time-slice clades. As might be expected, these are highly similar after considering the sample size. Silhouettes and triangles represent the groups and are colored according to how they are defined in each of the two analyses. Inset*: Within Afrotheria, none of the groups identified by a time slice at 86 million years are large enough for further study (i) Afrosoricida, $N=37$; (ii) Macroscelidea, $N=14$; (iii) Tubulidentata, $N=1$; (iv) Paenungulata, $N=11$. However, as a group in its own right, Afrotheria is large enough to study ($N=63$). We also highlight the three small clades within Paenungulata that for some reason are studied separately by Poe et al. (2020) by the orange shapes both on the tree and on the corresponding silhouettes: (□) Hyracoidea, $N=4$; (◯) Sirenia, $N=4$, (•) Proboscidea, $N=3$. All sample sizes reported here are for the larger data set.

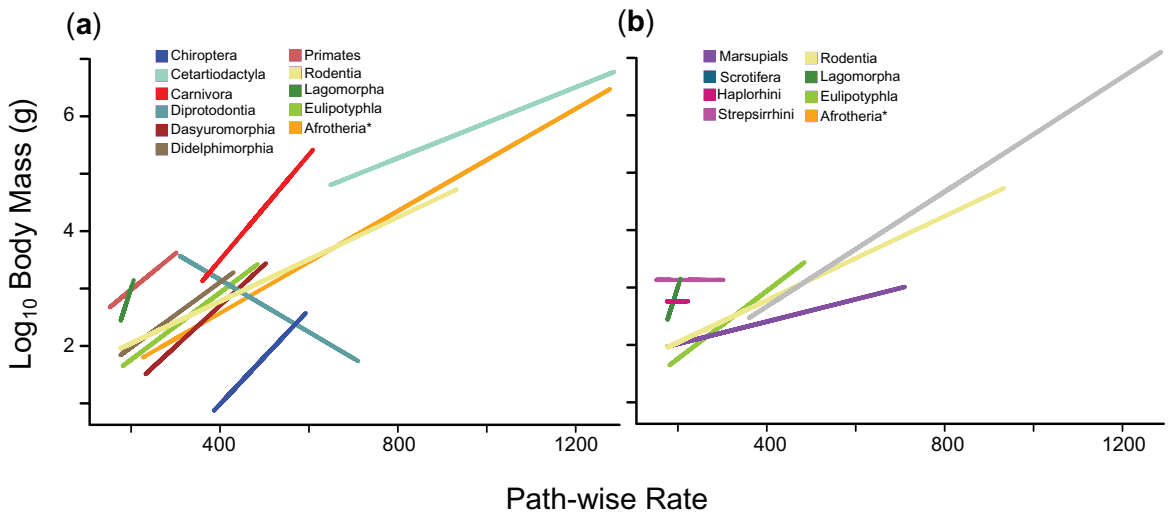

FIGURE 2. Results of phylogenetic generalized least-squares regression analyses which estimate the relationship between path-wise rate and body size in a) taxonomic groups and b) time-slice groups. Predicted slopes for each group are shown as estimated directly from the parameters in Table 1.

reason to devalue taxonomic clades over any alternative clade definition: both can provide useful information, as long as appropriate statistical approaches accounting for shared ancestry are used. Debates about clade definitions are likely to rage on, with little consensus to be found. Many named taxonomic groups have biological

TABLE 1. Parameter estimates from models estimating the relationship between body size and path-wise rate for (a) taxonomic groups and (b) time-slice groups.

(a)

Taxonomic[a]	Slope (β)	t	p
Afrotheria	0.0044	7.386	<0.001
Carnivora	0.0092	7.931	<0.001
Cetartiodactyla	0.0031	6.88	<0.001
Chiroptera	0.0082	8.269	<0.001
Dasyuromorphia	0.0071	8.968	<0.001
Didelphimorphia	0.0057	33.52	<0.001
Diprotodontia	−0.005	−6.044	<0.001
Eulipotyphla	0.0059	4.42	<0.001
Lagomorpha	0.0244	2.019	0.0436
Primates	0.006	2.537	0.0112
Rodentia	0.0037	10.742	<0.001

(b)

Time slice[b]	Slope (β)	t	p
Scrotifera	0.005	13.477	<0.001
Marsupials	0.00199	3.749	<0.001
Eulipotyphla	0.0059	4.309	<0.001
Lagomorpha	0.0244	1.982	0.0475
Haplorrhini	0.0054	1.865	0.0623
Strepsirrhini	0.0091	1.533	0.1254
Rodentia	0.0037	10.35	<0.001

[a]These values are not identical to those reported in Baker et al. (2015) because here we do not separate aquatic species in order to facilitate direct comparisons. Results are qualitatively identical if an aquatic dummy code is included in the model.
[b]These slopes may appear different to those plotted by Poe et al. (2020) owing to their use of inappropriate models (see Materials and Methods).

meaning, by way of shared derived traits and shared evolutionary histories. We suggest that unlike the rather disheartening suggestion to abandon the use of clades in our studies of evolutionary patterns and processes we can instead continue to use the phylogenetic "*corrections*" that are already the standard for most comparative evolutionary analysis.

MATERIALS AND METHODS

Body Size Data Set

Here, we use the data set used by Baker et al. (2015) ($N=3321$) which collated body size data from two sources (Ernest 2003; Jones et al. 2009) and matched it to a recent mammalian phylogeny (Fritz et al. 2009). Although Poe et al. (2020) use a larger data set of mammals ($N=3845$), the increase in number is attributable to the use of >300 species where body mass has been extrapolated using (nonphylogenetic) regressions from other mammals (Jones et al. 2009). These species were explicitly excluded both here and by Baker et al. (2015) owing to the fact that such data are highly biased and noninformative for phylogenetic analysis. However, we also reach the same conclusions if we repeat all analyses using this larger (potentially biased) data set with one minor exception—the slope for lagomorphs is marginally nonsignificant ($p=0.06$). This is likely attributable to minimal variation in the path-wise rate for this group (they span less than 2% the total variation in path length).

Clade Definitions

We assigned mammals to taxonomic clades as in Bininda-Emonds et al. (2007) and time-slice clades using a time slice of 86 MYA (Poe et al. 2020, Fig. 2). We highlight that applying a time slice of 86 MYA results in a total of 15 unique clades (Fig. 2) and a slightly different division of Afrotherian mammals to that presented by Poe et al. (2020) (Fig. 2, inset).

Estimating Trends in Body Size Evolution

We estimated two phylogenetic generalized least-squares regressions determining the relationship between body size and path-wise rate within (i) taxonomic clades and (ii) time-slice clades. For each analysis, we used standard contrast coding and interactions allowing the slope to vary between clades. Importantly, although Poe et al. (2020) chose to study three small groups ($N=3$ and $N=4$), to "*increase sample size*", we apply the original sample size criteria for inclusion in our regression analysis (but results are no different if we include them)—ensuring at least 10 data points per parameter estimated (Freckleton and Watkinson 2001). Ensuring a large enough sample size for analysis is critical as it means any conclusion drawn from the data is statistically meaningful. With this in mind, it is important to recognize that our analyses—and those of Baker et al. (2015)—still include all species and their phylogenetic relationships—we simply do not estimate slopes for groups with too few data.

Using each model (the parameters of which are presented in Table 1), we then inferred the ancestral body size at each internal node of the phylogeny using the methods described in Baker et al. (2015). This allowed us to study whether there is a general tendency towards body size increase by assessing the evolutionary change in body size along individual branches of the phylogenetic tree. We then used an exact binomial test to test whether there is a significant number of decreases compared to decreases. This is a robust, across-lineage, phylogenetically informed test.

Finally, neither here nor in the main body of this article do we quantitatively compare our results to those reported by Poe et al. (2020). Although our slopes visually are very similar (Fig. 2 and their Fig. 4), Poe et al. (2020) provide no parameter estimates. Furthermore, Poe et al. (2020) "*perform tree-wide analyses on multiple separate subtrees*" and—unlike Baker et al. (2015) and

our analyses herein—studied clades independently. This includes the sign tests reported by (Poe et al. 2020) that ascertain whether there are more positive slopes than negative slopes. Consequently, all of their results are biased owing to the nonindependence issues described above and in their original paper. Finally, we find it impossible to recover the reported negative slope for Eulipotyphla (see Poe et al. 2020, their Fig. 4) unless we use nonphylogenetic regressions. This, ironically, is a violation of the assumption of nonindependence far worse that not accounting for the shared ancestry among clades: the effect of this is so severe that the spurious negative slope is actually positive after accounting for phylogeny (see Eulipotyphla, Fig. 2).

Calculation of Bayes Factors

One way we could ascertain if there is statistical support that one clade definition fits the data better than another is to use standard goodness-of-fit metrics. Here, we use Bayes Factors (BF) to determine which model is preferred, calculated as $BF = -2\log_e[m_{\text{taxonomic}}/m_{\text{timeslice}}]$, comparing the marginal likelihood of a model that estimates slopes for taxonomic groups ($m_{\text{taxonomic}}$) to one estimating slopes for time-slice groups. To do this, we re-ran our maximum likelihood regression analyses in a Bayesian Markov Chain Monte Carlo framework implemented in BayesTraits (Pagel et al. 2004). In each model, we estimated marginal likelihoods using stepping-stone sampling (250,000 iterations for each of 500 stones) drawing values from a beta distribution $\alpha = 0.4\beta = 1$ (Xie et al. 2010). The Bayesian analyses provide results that are qualitatively identical to the maximum likelihood models in both the original paper (Baker et al. 2015) and the present analysis.

Data Availability Statement

No new data were generated or analyzed in support of this research.

Funding

JB was funded by the Leverhulme Trust (ECF-2017-22). CV was funded by a Leverhulme Trust Research Project Grant (RPG-2017-071). MP is funded by a Leverhulme Trust Research Project Grant (RPG-2019-170).

References

Baker J., Meade A., Pagel M., Venditti C. 2015. Adaptive evolution toward larger size in mammals. Proc. Natl. Acad. Sci. USA 112:5093–5098.

Barraclough T.G., Humphreys A.M. 2015. The evolutionary reality of species and higher taxa in plants: a survey of post-modern opinion and evidence. New Phytol. 207:291–296.

Bininda-Emonds, O.R., Cardillo, M., Jones, K.E., MacPhee, R.D., Beck, R.M., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L. and Purvis, A., 2007. The delayed rise of present-day mammals. Nature 446:pp.507–512.

Cantino, Philip D., and Kevin De Queiroz, eds. PhyloCode: a phylogenetic code of biological nomenclature. CRC Press, 2020.

Ernest S.K.M. 2003. Life history characteristics of placental nonvolant mammals. Ecology 84:3402.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Forey P.L. 2002. PhyloCode pain, no gain. Taxon 51:43–54.

Freckleton R.P., Watkinson A.R. 2001. Nonmanipulative determination of plant community dynamics. Trends Ecol. Evol. 16:301–307.

Fritz S.A., Bininda-Emonds O.R.P., Purvis A. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. Ecol. Lett. 12:538–549.

González-del-Pliego P., Freckleton R.P., Edwards D.P., Koo M.S., Scheffers B.R., Pyron R.A. et al. 2019. Phylogenetic and trait-based prediction of extinction risk for data-deficient amphibians. Curr. Biol. 29:1557–1563. e1553.

Hadfield J.D., Nakagawa S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J. Evol. Biol. 23:494–508.

Humphreys A.M., Barraclough T.G. 2014. The evolutionary reality of higher taxa in mammals. Proc. R Soc. Lond. B Biol. Sci. 281: 20132750.

Jones K.E., Bielby J., Cardillo M., Fritz S.A., O'Dell J., Orme C.D.L. et al. 2009. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology 90:2648.

Martins E.P., Hansen T.F. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.

Pagel M. 1997. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Pagel M., Meade A., Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. Syst. Biol. 53:673–684.

Poe S., Anderson C., Barnett J. 2020. On the selection and analysis of clades in comparative evolutionary studies. Syst. Biol.

Raftery A.E. 1996. Hypothesis testing and model selection. In: Gilks WR, Richardson S., Spiegelhalter D.J., editors. Markov chain Monte Carlo in practice. London, UK: Chapman & Hall. p. 63–187.

Sánchez-Reyes L.L., Morlon H., Magallón S. 2017. Uncovering higher-taxon diversification dynamics from clade age and species-richness data. Syst. Biol. 66:367–378.

Venditti C., Meade A., Pagel M. 2011. Multiple routes to mammalian diversity. Nature 479:393–396.

Waddell P.J., Okada N., Hasegawa, M. 1999. Towards resolving the interordinal relationships of placental mammals. Syst. Biol. 48:1–5.

Xie W., Lewis P.O., Fan Y., Kuo L., Chen M.-H. 2010. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60:150–160.