Comparing *cal3* and other a posteriori time-scaling approaches in a case study with the pterocephaliid trilobites

David W. Bapst and Melanie J. Hopkins

Abstract.—Reconstructing the tree of life involves more than identifying relationships among lineages; it also entails accurately estimating when lineages diverged. Paleontologists typically scale cladograms to time a posteriori by direct reference to first appearances of taxa in the stratigraphic record. Some approaches use probabilistic models of branching, extinction, and sampling processes to date samples of trees, such as the recently developed *cal3* method, which stochastically draws divergence dates given a set of rates for those processes. However, these models require estimates of the rates of those processes, which may be hard to obtain, particularly for sampling. Here, we contrast the use of *cal3* and other a posteriori time-scaling approaches by examining a previous study that documented a decelerating rate of morphological evolution in pterocephaliid trilobites. Although aspects of the data set make estimation of branching, extinction, and sampling rates difficult, we use a multifaceted approach to calculate and evaluate the rate estimates needed for applying *cal3*. In agreement with previous simulation studies, we find that the choice of phylogenetic dating method impacts downstream macroevolutionary conclusions. We also find contradictory evolutionary inferences between analyses on ancestor–descendant contrasts (based on ancestor trait reconstruction methods) and maximum-likelihood parameter estimates. Ancestral taxon inference in *cal3* corroborates previously hypothesized ancestor–descendant sequences, but *cal3* suggests greater support for budding cladogenesis than anagenesis. This case study demonstrates the potential and wide applicability of the *cal3* method and the benefits afforded by choosing *cal3* over simpler a posteriori time-scaling approaches.

David W. Bapst. Department of Earth & Planetary Sciences, University of California, Davis, One Shields Avenue, Davis, California 95616-8605, U.S.A., and South Dakota School of Mines and Technology, 501 E. St. Joseph, Rapid City, South Dakota 57701, U.S.A. E-mail: dwbapst@gmail.com
Melanie J. Hopkins. Division of Paleontology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, U.S.A. E-mail: mhopkins@amnh.org

Introduction

Phylogenies of fossil taxa are growing in both size and number. As a result, they are becoming a common framework for understanding the evolutionary patterns of ancient life. Macroevolutionary analyses often require dated phylogenies, but undated cladograms are much more common in the paleontological literature, in part because it is difficult to determine accurate divergence times from the fossil record alone. Furthermore, recent simulation analysis of commonly used dating approaches for paleontological phylogenies found that some methods had large effects on the precision and accuracy of downstream macroevolutionary analyses (Bapst 2014).

In this paper, we distinguish between two types of dating approaches that have become prevalent in the very recent literature. Fossil,

"tip-dating" approaches are a specific family of probabilistic approaches that simultaneously infer both relationships and divergence dates for a set of taxa that are at least partially from the fossil record (Pyron 2011; Ronquist et al. 2012). These analyses are usually done with Bayesian phylogenetics software, such as MrBayes or BEAST 2, both of which allow for taxa to be placed as sampled ancestors (Gavryushkina et al. 2014; Zhang et al. 2016). Tip-dating analysis requires both character and stratigraphic information data, implementation of a model of clocklike morphological change, and a probability model that describes the expected waiting times between branching events, such as variants of the birth-death-sequential-sampling (BDSS) model (Stadler 2010; Stadler and Yang 2013). While tip-dating methods can be applied to data sets containing only fossil taxa (Lee et al. 2014;

0094-8373/17

TABLE 1.   Commonly used a posteriori time-scaling (APT) approaches in paleontology.

Methods	Primary references	Description
Minimum-node dating (MND, basic)	Norell 1992; Smith 1994	Assigns branching node ages as the first appearance time of the earliest appearing taxon descended from that node. Incongruence between the phylogeny and the order of stratigraphic appearance results in multiple successive internode edges of zero length.
Minimum-length branches	Laurin 2004	Scales all branches of a MND tree so they are greater than or equal to an independently and arbitrarily defined value, while preserving the relative distance between the terminal tips and the root node.
Equal branch-length sharing	Ruta et al. 2006; Brusatte et al. 2008; Bell and Lloyd 2015	Shifts root age of an MND tree earlier in time by some constant and then equally reapportions the duration of non–zero length internode edges along successive zero-length child edges.
Ruta branch-length sharing	Ruta et al. 2006	Shifts root age of an MND tree earlier in time by some constant and then uses number of inferred character changes as weights to reapportion the duration of non–zero length internode edges along successive zero-length child edges.
cal3	Bapst 2013a	Stochastic sampling of node ages relative to their likelihood under a birth-death-sampling model calibrated via branching, extinction, and sampling rates given as input. These a priori rates are typically estimated from occurrence or range data.

Bapst et al. 2016), other methods (sometimes referred to as "fossilized birth–death" methods, although this name technically refers only to the model employed), use fossil occurrences without character data, loosely constrained in their phylogenetic placement, to date molecular phylogenies (Heath et al. 2014), and thus cannot be applied currently to fossil-only data sets.

This is in contrast to a posteriori time-scaling (APT) approaches, which date a preexisting unscaled topology (i.e., a cladogram) given a set of stratigraphic data for the taxa involved. This allows for stratigraphy-free approaches to building cladograms, and can be applied to cases in which usable character data are not known for all included taxa, such as when topologies are created through supertree approaches or by combining taxonomic and phylogenetic data. Most APT approaches are simple algorithms; the most common method assigns minimum-node ages, wherein each divergence is as old as the first occurrence of the oldest descendant taxon (Table 1). These minimum node–dated (MND; also sometimes referred to as basic time-scaling) trees often contain a large number of very small or zero-length branches that may be biologically implausible and often violate the mathematical assumptions of many phylogeny-based macro-evolutionary analyses (Hunt and Carrano 2010; Bapst 2014).

To avoid this, various ad hoc modifications have been invented to alter the branch lengths to have some positive value (Table 1) by removing zero-length branches. Adjustments are often relative to some given constant, the nature of which differs among methods. One frequently used approach involves setting a minimum branch length and shifting divergences backward in time so that all edges (internode branches) attain this minimum length (Laurin 2004; Bapst 2014). A second, equally popular approach shifts the root age backward some set amount and then reapportions the duration of positive-length edge along successive zero-length child edges. This latter approach is sometimes applied so that the reapportioned time is shared equally along branches or relative to the number of inferred character changes, as if under a strict morphological clock (Ruta et al. 2006; Brusatte et al. 2008; Bell and Lloyd 2015). Both of these approaches are problematic, mainly for their reliance on the user-supplied ad hoc constants (the minimum branch length, the additional time added to the root age), without any direct reference to underlying processes.

Not all APT algorithms are simple. Tomiya (2013) stochastically sampled node ages for a tree of fossil taxa constrained by several molecular dating estimates, resulting in a sample of trees that reflect a uniform uncertainty in node ages. However, many data sets lack molecular clock constraints, so

this approach is not widely applicable. Furthermore, the uncertainty of clade ages is expected to differ relative to the degree of sampling in the fossil record and the pace of taxonomic turnover. Bapst (2013a) introduced a probabilistic three-rate calibrated time-scaling method known as *cal3*, which requires a priori information about the rates of branching, extinction, and sampling to calculate probability densities of the uncertainty in divergence dates, under a variant of the BDSS model (Stadler 2010). These distributions could then be stochastically sampled to produce samples of trees that reflect the uncertainty in node ages under this model, similar to Tomiya's approach. In contrast to earlier APT approaches, *cal3* relies on a probabilistic model that makes explicit reference to the processes of diversification and sampling involved. Bapst (2014) compared several APT approaches, including *cal3*, in their effect on a number of typically applied analyses that attempt to reconstruct various evolutionary patterns using time-scaled trees of fossil taxa. In general, *cal3* always performed as well as or better than other APT methods. For some comparative analyses, such as fitting models of trait evolution, all APT methods showed some bias toward an incorrect inference, although *cal3* showed the least bias relative to other methods.

There are also a number of reasons why *cal3* may be preferable to tip-dating and similar methods, at least in their current formulation. Most notably, current tip-dating methods can only treat operational taxon units as point occurrences in time, ignoring that many fossil taxa (including fossil species) are persistent lineages observed over long intervals of geologic time. This sometimes results in tip-dating analyses arbitrarily choosing the first appearance time of fossil taxa or the midpoint of their duration as the date assigned to a taxon (Heath et al. 2014), a practice that may negatively impact the assumed sampling model (Bapst et al. 2016). This excludes the ability of the tip-dating methods to distinguish between anagenetic and budding cladogenetic modes of ancestor–descendant relationships, but which is a noted capability of *cal3*. Development of models of morphological evolution (as typically used in tip-dating) have also lagged behind that of molecular evolution

(but see Wright et al. 2016), and the Markov models currently employed may be inadequate for explaining observations of "static" fossil lineages with dozens of characters that exhibit no apparent change over millions of years (Wagner and Marcot 2010). Furthermore, *cal3* also has the capability, by allowing rates to be input for each individual taxon, to consider a wider range of clade-specific and time-heterogeneous rates than many tip-dating approaches.

However, the need for the three rate estimates has been viewed as a potential obstacle to applying the *cal3* method to empirical data sets and has led some workers to explicitly argue for the use of simpler APT approaches (e.g., Soul and Friedman 2015; Bokma et al. 2016; Brocklehurst 2016). This view is partly evidenced by the adoption of *cal3* by only a small number of studies to date (Hopkins and Smith 2015; Halliday and Goswami, 2016a,b; Puttick et al. 2016). In this paper, we address the barriers encountered by those who want to apply *cal3* to their data set. First, we clarify certain philosophical aspects of *cal3*'s approach, contrasted against other time-scaling methods. Then we describe several routes for obtaining rate estimates and address several typical misunderstandings of sampling rate models in paleobiology. We demonstrate rate estimation and the application of *cal3* to an empirical data set of Cambrian pterocephaliid trilobites (Hopkins 2011b, 2013b). We chose this data set because it is not amenable to standard procedures for estimating sampling rates, and thus the acquisition of sampling rates is not a straightforward procedure. We compare the resulting time-scaled trees to trees created using alternative APT approaches, particularly by comparing the results of phylogenetic comparative methods used in previous literature (Hopkins 2013b). In addition, we test whether the ancestor–descendant pairs inferred under *cal3* with sampled ancestors match previous systematic assessments (Hopkins 2011b).

Obtaining Rates for *cal3*

The *cal3* time-scaling method relies on a family of models sometimes referred to as the

fossilized birth–death or BDSS models (Foote 2001; Stadler 2010; Stadler and Yang 2013; Gavryushkina et al. 2014; Heath et al. 2014; Bapst et al. 2016; Zhang et al. 2016). Under BDSS models, processes of branching, extinction, and sampling of lineages are treated as Poisson processes, with events occurring relative to some instantaneous rates, granting several expected properties. The waiting times between events under time-homogenous Poisson processes are exponentially distributed, and the reciprocal of the rate is the expected waiting time to an event. As the exponential distribution is memory-less from any given starting point, the probability of a waiting time of a particular duration remains the same, regardless of when you start waiting for an event. Certain types of fossil data, such as taxonomic durations, result from an interaction of processes. For example, under a model in which species originate via budding cladogenesis (Foote 1996), observed taxonomic durations are clearly a function of both extinction rate (as waiting times from a birth event to a death event) and sampling, because the observed duration of taxa is dependent on the first and last sampling events of a taxon, not the true birth and death events (Foote and Raup 1996; Foote 1997).

Previously published literature may be invaluable for obtaining the rates necessary for *cal3* or providing a check on newly obtained rate estimates. The BDSS model is a generalization of the birth–death model (Kendall 1948), commonly applied in both paleobiology (e.g., Raup 1985; Foote 1996, 1997, 2000, 2001) and evolutionary biology (Nee et al. 1992; Nee 2006). Thus, estimates of birth and death rates are often found in the literature, often as the instantaneous "per capita" origination and extinction rates developed by Foote (2000). These estimates of the branching and extinction rates are appropriate for use with the *cal3* time-scaling approach, presuming that they were calculated at the same taxonomic scale as the taxon units on the provided cladogram.

While it is common to model sampling in the fossil record as a Poisson process (Strauss and Sadler 1989; Foote 1997; Starrfelt and Liow 2016), absolute parametric measures of sampling intensity, like sampling rate, are infrequently reported. This may stem from some uncertainty as to what the sampling process represents. In most descriptions of models in which sampling is a Poisson process, sampling encapsulates all processes involved in whether an individual that was alive at a particular moment in time is observed in the rock record: burial, preservation, survival of the material to the modern, and its collection and identification today. In modeling sampling as a Poisson process, sampling events are assumed to be rare, discrete events, implying that sampling events reflect observations from collections taken from different stratigraphic horizons (i.e., of different ages).

The most direct estimates of sampling rate were developed for data sets of fossil occurrences. For example, Solow and Smith's (1997) horizon-based estimate only requires the number of occurrences (i.e., the number of stratigraphic horizons at which a taxon has occurred) and the precise duration of each taxon to obtain the maximum-likelihood estimate of the instantaneous sampling rate. However, both of these pieces of information are unknown for many data sets. Alternative methods work on indirect relationships between sampling rate and other types of observable data, although this risks the effect of confounding variables. One popular set of approaches infers sampling and extinction rates from the frequency distribution of taxon durations across time intervals (Foote and Raup 1996; Foote 1997), while an alternative set of methods relies on reverse and forward survivorship curves to estimate origination, extinction, and sampling rates (Foote 2001). However, these methods may be sensitive to deviation from a Poisson process or to biases in the estimated number of taxa known from a single occurrence or interval. These biases can sometimes be overcome by expanding the number of taxa considered, especially if short-lived taxa or taxa known from single collections were excluded (which may be the case, as these taxa often have poor morphological data).

Analyses of discrete interval durations (e.g., Foote and Raup 1996) typically provide per interval sampling probabilities, but it is sometimes possible to convert these to

sampling rates. Per interval sampling probability is not independent of extinction rate (Foote 1997), but assuming that taxon durations are typically longer than the time-interval durations, probabilities can be converted to instantaneous sampling rates using the following equation (following from eq. 26 of Foote 2000):

$$sampling\ rate = -\frac{log\ (1-sampling\ probability)}{interval\ length} \quad (1)$$

Note that in groups with high rates of turnover relative to interval length, such that taxon ranges typically do not extend through intervals, this conversion may underestimate the sampling rate. However, when appropriate, this conversion is useful, as sampling probabilities are more commonly reported than sampling rates (e.g., Foote and Sepkoski 1999; Wagner and Marcot 2013), and has allowed us to provide a number of published estimates of sampling rate (Table 2). Unfortunately, relative sampling intensity is more commonly measured via proxies, such as the number of geologic formations or the number of collections made over a time period (e.g., Peters and Foote 2001), which cannot be translated to sampling rates.

Silvestro et al. (2014) introduced a Bayesian method called PyRate that calculates posterior estimates of birth, death, and sampling rates from fossil occurrence data by stochastically reconstructing the original taxon durations. The default sampling model in PyRate is not time homogenous, which contradicts some assumptions of the model used by *cal3* (Bapst 2013a), but an alternative time-homogenous model can also be fit.

Considerations for Providing Rates for *cal3*

The true branching, extinction, and sampling processes are probably not like simple Poisson processes, particularly in that they are not homogeneous. These processes likely vary in complex ways across clades, time intervals, biogeographic regions, and depositional environments, probably even varying across the duration of single species in the fossil record (Holland 2003, 2016). For example, some of the best-sampled fossil groups can

have exceptionally long ghost branches. Among the well-sampled planktonic graptolites, the late Ordovician *Sinoretiograptus* appears to be most closely related to early glossograptids (Vandenberg 2003), implying a 15–20 Myr ghost branch. Under a homogenous Poisson process of sampling, it would be extremely unlikely to observe such a long gap in sampling of a lineage and *simultaneously* observe the high sampling frequency of other graptoloid taxa in the Ordovician. This suggests some environmental or spatial filtering of the rock record, with some regions or habitats not captured by the fossil record, and thus is a violation of the simple sampling model.

However, such violations do not immediately preclude the use of a model in some analysis. All statistical models are likely violated to some extent, but simple models can sometimes be robust to such violations, and thus the challenge is determining whether those violations have any perceptible effect on the resulting analyses. This has been rarely explored in the context of models of sampling processes in the fossil record, with one exception. Wagner and Marcot (2013) found that sampling models that allowed for heterogeneity across a data set of Cenozoic mammalian taxa estimated very different mean sampling probabilities (erroneously referred to as a "rate" in their paper) relative to the probabilities calculated by homogenous models. However, it is unclear whether this is a general result, and the degree to which sampling probability varies across taxa in real data sets is largely unexplored. If rate heterogeneity is a significant concern, between-taxon variation can be accounted for with *cal3*, which treats branching, extinction, and sampling rates on a per taxon basis if these rates are individually supplied for each taxon unit in input (Bapst 2013a). More often, the true extent of rate heterogeneity is unknown, and it should be noted that future work is needed to address how sensitive *cal3* is to such heterogeneity, or even just analytical error, in the input rate estimates.

Before applying the relaxed version, it is important to keep in mind that per unit (e.g., per taxon or per time interval) point

TABLE 2. Calculated values of instantaneous sampling rates from the literature, for various taxonomic groups, rankings, geographic regions, and geologic intervals. The majority of these estimates were reported originally as the per interval sampling probability of a single taxon, and were translated into instantaneous sampling rates on a million year timescale using equations from Foote (1997). The Ursidae sampling rate from Gavryushkina et al. (2014) was reported as the expected sampling proportion and was translated to an instantaneous sampling rate using the equations in Heath et al. (2014). Note that some papers, such as Foote and Sepkoski (1999) and Friedman and Brazeau (2011) reported multiple estimates (or ranges of estimates) based on different data treatments.

Taxon	Taxonomic level	Geographic region	Geologic Interval	Instantaneous sampling rate (per LMyr)	Reference
Anthozoa	Genera	Global	Phanerozoic	0.09, 0.12	Foote and Sepkoski 1999
Asterozoa	Genera	Global	Phanerozoic	0.05, 0.05	Foote and Sepkoski 1999
Bivalvia	Genera	Global	Phanerozoic	0.11, 0.13	Foote and Sepkoski 1999
Bivalvia	Species	Europe	Jurassic	0.41	Foote and Raup 1996
Blastozoa	Genera	Global	Phanerozoic	0.10, 0.12	Foote and Sepkoski 1999
Brachiopoda	Genera	Global	Phanerozoic	0.48, 0.87	Foote and Sepkoski 1999
Bryozoa	Genera	Global	Phanerozoic	0.21, 0.24	Foote and Sepkoski 1999
Cephalopoda	Genera	Global	Phanerozoic	0.33, 0.44	Foote and Sepkoski 1999
Chiroptera (Bats)	Genera	Global	Cenozoic	0.01	Eiting and Gunnell 2009
Chondrichtyes	Genera	Global	Phanerozoic	0.01, 0.02	Foote and Sepkoski 1999
Conodonta	Genera	Global	Phanerozoic	0.20, 0.43	Foote and Sepkoski 1999
Crinoidea	Genera	Global	Early Paleozoic	0.10	Foote and Raup 1996
Crinoidea	Genera	Global	Phanerozoic	0.08, 0.08	Foote and Sepkoski 1999
Echinodea	Genera	Global	Phanerozoic	0.16, 0.20	Foote and Sepkoski 1999
Gastropoda	Genera	Global	Phanerozoic	0.10, 0.15	Foote and Sepkoski 1999
Graptolithina	Genera	Global	Phanerozoic	0.18, 0.36	Foote and Sepkoski 1999
Macroperforate Planktonic Foraminifera	Species	Global	Cenozoic	1.66	Ezard et al. 2011
Malacostraca	Genera	Global	Phanerozoic	0.04, 0.08	Foote and Sepkoski 1999
Mammalia	Genera	Spain and Portugal	Neogene	1.05	Alba et al. 2001
Mammalia	Species	North America	Cenozoic	0.41	Foote and Raup 1996
Mammalia	Species	North America	Late Cretaceous	0.03, 0.04	Foote et al. 1999
Mammalia	Species	Spain and Portugal	Neogene	0.80	Alba et al. 2001
Osteichtyes	Genera	Global	Phanerozoic	0.03, 0.07	Foote and Sepkoski 1999
Ostracoda	Genera	Global	Phanerozoic	0.13, 0.15	Foote and Sepkoski 1999
Polychaeta	Genera	Global	Phanerozoic	0.01, 0.01	Foote and Sepkoski 1999
Porifera	Genera	Global	Phanerozoic	0.09, 0.12	Foote and Sepkoski 1999
Tetrapoda	Genera	Global	Devonian	0.044–0.18, 0.044–0.099, 0.042–0.085	Friedman and Brazeau 2011
Trilobita	Genera	Global	Phanerozoic	0.19, 0.39	Foote and Sepkoski 1999
Trilobita	Species	Single locality in Oklahoma	Upper Cambrian and Lower Ordovician	17.46	Foote and Raup 1996
Ursidae	Species	Global	Cenozoic	0.094	Gavryushkina et al. 2014

estimates, such as maximum-likelihood estimators, may exaggerate the true heterogeneity in a system. Paleontologists largely agree that there is real variation in branching and extinction rates (between clades, across time, at mass extinction events, etc.), but it has long been recognized that simulations of time-constant birth–death processes can easily mimic many of the minor radiations and extinctions observed in the fossil record (Raup et al. 1973), and thus it is difficult in practice to separate out true variation in these rates from

stochastic noise. Furthermore, some studies have suggested that long-term averages of origination and extinction rates are roughly equal (Stanley 1979; Sepkoski 1998) for both extinct and extant groups. This has motivated the development of methods that move beyond per unit point estimates of origination and extinction rates (e.g., PyRate; Silvestro et al. 2014). In cases in which the clade of interest is entirely extinct, this apparent equivalency could be artifactual: if point estimates are calculated for the long-term

average birth and death rates for the entire history of an extinct clade, the birth rate must equal the death rate, as no more taxa could have originated than go extinct (or vice versa). Ironically, the upside of this is that it may be reasonable to assume that branching and extinction rates are approximately equal for a given fossil data set and to use one as a proxy for the other for extinct clades in *cal3* (e.g., Bapst 2014).

Another issue with calculating rates under the BDSS process is that many paleontological data sets do not include every single taxon known from a given group, as many named taxa may be precluded due to incomplete data (e.g., stratigraphic, morphological, etc). Furthermore, some "observed" taxa may go unrecognized, particularly if there are too few complete specimens for systematists to recognize them as novel morphological species (perhaps a paleontological parallel to the incipient speciation problem; Etienne and Rosindell 2012). This may result in an underestimate of how many species are known from a small number of collections, and this could lead to inflated or deflated estimates of extinction and sampling rates, depending on the nature of the missed occurrences. Workers need to be cognizant of such potential biases and should optimally include as many taxa as possible when calculating rates.

Regardless of the complexities of measuring them, these three instantaneous per lineage rates of branching, extinction, and sampling required by *cal3* are grounded in a mechanistic worldview, in which observable phenomena like the appearance of groups in the fossil record are the product of underlying processes of diversification and sampling. While direct measurement of the required rates may not always be tractable with every data set, the literature is replete with previous estimates for those rates. On the contrary, there are no previous estimates for "minimum branch length," and our intuition could be a biasing factor for adjusting root ages, given the frequent disagreements between the expert opinions of paleontologists and molecular clock estimates for the timing of divergences. Thus, in the absence of measurements for the rates, the ideal alternative may be to apply

educated guesswork to identify a probable range of rates for *cal3*, rather than using an alternative APT approach that relies on ad hoc scaling constants.

Empirical Case Study: Methods

Data

We applied *cal3* and other APT methods to a data set of pterocephaliid trilobites that was the basis of previous studies from Hopkins (2011a,b, 2013a,b). The cladogram dated in the present study with APT approaches is the single most parsimonious tree of 42 taxa returned when continuous characters were gap-weighted and is taken from Hopkins (2011b). Twelve taxa that do not have stratigraphic occurrence data or landmark data were dropped from this cladogram. Trait data used in phylogenetic comparative analyses in this study are principal component (PC) scores from an ordination of superimposed landmark data summarizing the shape of the trilobite cranidia (Hopkins 2013a).

Temporal data were extracted from a composite section of Laurentian-wide occurrences of Steptoean trilobite taxa produced using constrained optimization (CONOP9; Sadler et al. 2003) and described in previous papers (Hopkins 2011b, 2013b). CONOP approaches the correlation problem in two steps: first, as a seriation problem (the ordinal position of each taxon's first and last appearance), and second as a scaling problem (the relative position within the section of each taxon's first and last appearance). It is during the first step that the quality of the solution is assessed. Shifts in the position of the first and last appearances differentially affect the quality of the solution depending on the number of co-occurrences with other taxa at their range ends. It is common for there to be multiple equally well-fit solutions to the correlation problem, and from these, it is possible to determine a range of best-fit positions given that set of equally well-fit solutions (e.g., Cody et al. 2008). Hopkins (2011b, 2013b) used a single set of ranges consisting of the mean first and last appearances from 100 discovered solutions, but this ignores the nonindependence of the

uncertainties across solutions. Thus, in this study, we treat each of the 100 discovered solutions of Hopkins (2011b) as an inseparable ensemble of estimated first and last appearances. Each composite section was converted to millions of years by scaling to the first appearance of *Glyptagnostus reticulatus* and *Irvingella major*, which represent the base of the Steptoean (~497 Ma) and the base of the overlying Sunwaptan Stage (~473 Ma), respectively (Palmer 1962; Ludvigsen and Westrop 1985; Gradstein et al. 2012). This scaling effectively assumes a constant rate of "deposition" (roughly, 1 m per 0.03 Myr), but this is reasonable, because the relative position of each taxon in each composite section was estimated using the scaling option in CONOP that best approximates actual time, particularly over intervals when rapid origination or extinction took place (Sadler 2007).

In addition to the stratigraphic ranges themselves, which are output from CONOP as first and last occurrence dates in each composite, we attempted to count the number of collections in which each species occurs. This is obfuscated somewhat in the original data input to CONOP, as CONOP only requires the stratigraphic heights of the horizons for the first and last appearances of a taxon at each site, even though many species have been sampled at horizons in between the first and last at some localities (e.g., see stratigraphic columns in Palmer 1965). Thus, we calculated the minimum number of collections by counting each locality a taxon was found at as two collections. Taxon occurrences from a single horizon at a particular site (i.e., the first and last horizon are identical) were counted as only a single collection.

Calculating Sampling and Diversification Rates for *cal3*

A typical approach to obtaining sampling estimates is to use the frequency distribution of stratigraphic ranges, whether in discrete intervals or continuous time, to fit models of extinction and incomplete sampling (Foote and Raup 1996). We used the maximum-likelihood method for continuous time ranges from Foote (1997), as implemented in the package

paleotree (Bapst 2012) in R (R Core Team 2016), to estimate the sampling and extinction rate from each of our 100 CONOP composites, resulting in samples of extinction and sampling rates. This method assumes that the range data is approximately exponentially distributed (ignoring taxa found at a single horizon; i.e., singletons), which is satisfied here (example in Fig. 1A). However, a sizable number of singletons is also expected by this method, but the CONOP composites we used show an apparent lack of true singletons, possibly due to the strategy employed by CONOP of minimally lengthening taxon durations to eliminate inconsistency in the order of taxa across different sections. Thus, we treated any taxon with a duration less than 0.1 Myr as a singleton for that composite (effectively, taxa whose range was less than ~3.1 m in that composite).

A different approach, introduced by Solow and Smith (1997), uses a maximum-likelihood estimator to calculate the instantaneous sampling rate based on taxon durations and the number of collections (sampling events) in which a taxon was found. As described above, we used the minimum number of collections to obtain estimates of the minimum sampling rate. This method is implemented in the R package paleotree.

Results presented for our calculations of sampling and extinction rates are based on pruned composites consisting only of the 30 pterocephaliid taxa on our pruned cladogram, as our main interest is the sampling and diversification rates of the taxa included in our phylogenetic data set. Thus, all analyses in this paper are for those taxa for which cladistic, stratigraphic, and morphometric data were available. Similar analyses were performed for the entire 163 trilobite species included in the CONOP composites, but rate estimates did not qualitatively differ (see Supplementary Material), suggesting that the selected 30 pterocephaliid species are an unbiased subsample of the larger taxonomic data set.

Applying APT Approaches

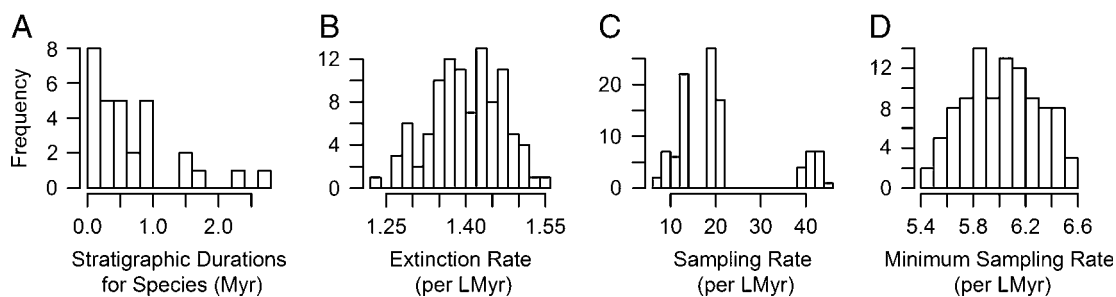Hopkins (2013b) presented phylogenetic comparative analyses using a phylogeny

FIGURE 1. Estimating sampling and extinction rates from the pruned occurrence data. Frequency distributions of stratigraphic durations from 100 individual CONOP solutions (example from one such run in A) were used to calculate samples of estimated extinction rates (B) and sampling rates (C) via maximum likelihood (Foote 1997). The number of first and last appearances of taxa across all sections used in the CONOP analysis provide a minimum number of occurrences, which were used with the duration data to calculate minimum bounds for sampling rate (D) across the 100 CONOP solutions, using a different maximum-likelihood method (Solow and Smith 1997).

scaled under the "equal" algorithm (Table 1), times of observation in units of stratigraphic height, and the root age offset by 1 m. For the absolute timescale used in this study, 1 m is effectively ~0.03 Myr. We expanded this to nine variations of a posteriori time-scaling: (1) a set of trees time-scaled using the minimum-node dating approach; (2–4) trees time-scaled using equal with the root age adjusted backward by 0.03 Myr (i.e., 1 meter), 0.1 Myr, or 0.2 Myr; (5–7) trees time-scaled with the minimum branch length algorithm (MBL), using respectively 0.03 Myr, 0.1 Myr, or 0.2 Myr as the MBL variable; and (8–9) *cal3* using the median sampling and extinction rate estimates calculated from the duration frequency analyses (see "Results"), and allowing and disallowing ancestor inference (Bapst 2013a). Each APT approach was applied once for each of the 100 CONOP composites, such that each approach is represented by 100 time-scaled trees. The age of the occurrence selected for each taxon (i.e., the times of observation) can have a significant impact on comparative methods (Bapst 2014). Our goal was to examine how different APT approaches impact the findings of Hopkins (2013b), so we followed the protocol used there of time-scaling with the first appearances as the times of observation. The morphological data set used is the mean shape of the cranidium, averaged across time and space for each species, and may be considered the expected shape of the first representatives of each species. All of these approaches were applied

using functions in the R package paleotree (Bapst 2012).

Phylogenetic Comparative Methods for Testing Time-Dependent Rates of Trait Evolution

Hopkins (2013b) recovered a negative trend through time in the rate of ancestor–descendant morphological change, using a multivariate approach. These rates were calculated via maximum-likelihood reconstructions of ancestral node values (Schluter et al. 1997; as implemented in R package ape [Paradis et al. 2004]) for all PCs, on a unit-length tree (a phylogeny in which all internode edge lengths are 1), and then dividing the Euclidean distance between the ancestor and descendant PC values for all edges by their respective edge lengths on the time-scaled tree. The correlation between "time" (technically, stratigraphic height) and rate was measured as Kendall's tau. We applied the exact same protocol for all samples of time-scaled trees, although with the additional step of ignoring any infinite rates estimated when edge lengths were zero (which are impossible under equal or MBL but not MND or *cal3*, which were not applied in the original study). Alternative analyses in which zero-length branches were not ignored did not produce qualitatively different results (see Supplementary Material for details).

To assess the degree to which similar procedures might produce different results, we conducted two additional tests of the

temporal trend in the rate of trait evolution. First, we applied the same ancestor–descendant divergence analysis described above, except with ancestral node values estimated using the time-scaled trees, not a unit-length tree. This avoids possible misestimation of ancestral values due to ignoring temporal structure.

Second, we applied recently developed methods for fitting models of trait evolution to multivariate data sets, provided in the R package mvMorph (Clavel et al. 2015). Following best practices recommended by Slater and Pennell (2014) for testing for time-dependent change in rates of morphological change, we fit a version of the "early-burst" (EB) model, relaxed to allow either an exponential decrease or increase with time of the rate of morphological change (so, allowing for both late-burst and early-burst scenarios), which is functionally identical to the "accelerating–decelerating" (ACDC) model (Blomberg et al. 2003). The support for an early-burst or late-burst scenario can be summarized using the maximum-likelihood value of the beta parameter returned by the function mvEB, but rather than examining this value in its raw state, we transformed it to the more interpretable number of rate half-lives elapsed over a clade's evolutionary history, which is $T/[\log(2)/\text{beta}]$ (Slater and Pennell 2014), where T is tree depth. Tree depth, however, varies within our samples of examined trees, so instead we used a flat 5 Myr as T (close to the total duration of the pterocephaliids in the fossil record), thus calculating the number of doublings (or halvings) per 5 Myr. Unfortunately, we could not perform this analysis on all 30 PCs, as this was prohibitively slow, so we instead limited our analyses to the first four PCs, which account for more than 95% of the variance in the data set, decreasing our computational time by almost three orders of magnitude. To avoid numerical issues related to zero-length branches in model-fitting comparative analyses, we slightly adjusted terminal edge lengths by 0.001 Myr, following Bapst (2014).

Testing Ancestor–Descendant Inference

We examined the pattern of ancestor–descendant inferences within an additional set of phylogenies using *cal3*, with last appearances as the times of observation for the taxon units (i.e., tips will reflect the last appearance times). This was done because the use of first appearances will lead to an underestimation of the frequency of inferred ancestors, as branching of child lineages from taxa later in their duration may be inappropriately downweighted by the *cal3* algorithm. While putative ancestors are tagged as such in the output of *cal3* (as well as by evolutionary mode; i.e., anagenesis or cladogenesis), descendants are not tagged as such, because *cal3* only recognizes one sister lineage as the descendant, but this may be a single taxon or a very large clade containing many taxa. The actual identity of the descendant in the latter case is somewhat amorphous, but the most probable identity is that of the earliest appearing taxon in that clade. We use this guideline to identify the descendant taxon for each inferred ancestor.

All data, input files and programming scripts for recreating all analyses and figures can be found at Dryad repository http://datadryad.org/handle/10255/dryad.122995.

Empirical Case Study: Results

Estimates of Sampling and Diversification Rates

As trilobites are a group of marine invertebrates with a rich fossil record, a number of previous estimates for sampling rate already exist in the literature, at a range of taxonomic, temporal, and geographic scales (Table 2). These estimates can have a correspondingly large degree of variability across those scales, with a global Phanerozoic-scale study of genera reporting 0.19–0.39 per lineage million years (per LMyr; Foote and Sepkoski 1999), while the sampling rate estimate from a single locality in Oklahoma has an apparent sampling rate of 17.46 per LMyr (Foote and Raup 1996). This last value is exceedingly high, suggesting that the trilobite record might be very complete at small spatial scales.

Our maximum-likelihood estimates of sampling rate showed considerable variation (more than 30 units per LMyr), especially compared with the much narrower

distribution of extinction rate estimates (Fig. 1B,C). The median sampling rate from our pruned data set was 19.07 per LMyr, and the median extinction rate was 1.40 per LMyr. The minimum sampling rate estimates (following Solow and Smith 1997) were much more constrained than the estimates from the range frequency distributions (Fig. 1D), with a median of 6.02 per LMyr.

Overall, these calculations suggest that the record of the pterocephaliid trilobites within the Great Basin is extremely complete and well-sampled, comparable to the sampling estimates from Foote and Raup (1996) based on a single section of upper Cambrian–Ordovician trilobites from Oklahoma (Table 2). Although the cross-validation of the high sampling rates estimated from the range frequencies versus the high minimum-rate estimates is reassuring, it is important to evaluate the realism of these rates based on estimates in other groups and their implications for taxonomic completeness and turnover. For example, the median sampling and extinction rate estimates would imply an average species duration of 0.71 Myr and an average waiting time between sampling events (collections) for a single species within its duration of 0.05 Myr, and given those values, that the average species will be collected at 13 horizons over its duration. Although we have only limited information on the number of collections a taxon was sampled at, the average minimum number of horizons a taxon occurred at was 5.8, which would be consistent with the above estimate. The mean average duration of a species in our pruned data set was 0.66 Myr, which also matches well with our parameter estimates. We can also use these rate estimates to calculate taxonomic completeness: the probability of sampling a taxon at least once, equivalent to the expected proportion of taxa that were observed (Foote and Raup 1996). The expected taxonomic completeness for this data set is 0.93, implying that the pterocephaliid record is exceedingly complete but that some fraction of species may still have been unsampled. Overall, the median sampling (19.07 per LMyr) and extinction (1.40 per LMyr) rate estimates imply nothing that seems unlikely a priori in comparison to

what we know of the pterocephaliid record and the fossil record of other groups, and thus we adopt these as valid estimates for use in *cal3*. We choose to use median values rather than sampling from our sample of rate estimates, as the resulting distributions are multimodal, with extreme outliers likely due to outlier sets of CONOP ranges. Additionally, as the pterocephaliids are extinct (and their entire evolutionary history is captured here, with taxa only omitted due to missing character or stratigraphic information), the long-term branching rate must equal the extinction rate, and thus we use the median extinction rate as an estimate of the branching rate.

Comparing Divergence Dates on Time-Scaled Trees

While there are a variety of ways for sets of time-scaled trees to differ (Bapst 2014), one simple way to visualize their effect on divergence dates is to calculate the distribution of node ages and visualize these on a common absolute timescale. We do this both for the root divergence of the tree (the split between *Aphelaspis haguei* and the other 29 pterocephaliid species) and a more derived node, in this case the split between the sister genera *Sigmocheilus* and *Pterocephalia* (Fig. 2). Little variation is seen among the root divergence dates, except that the minimum branch length algorithm shifts the root age back about 1 Myr, suggesting many of the edge lengths may be considerably less than 0.2 Myr. The slight variation among sets of trees time-scaled with equal matches the difference in their root age offset values. More variation is seen in the age estimates of the more nested node, with a linear difference in node ages observed as minimum branch length is increased. The three equal method treatments estimated the same divergence date for the *Sigmocheilus–Pterocephalia* split, which agrees with the description of the underlying algorithm (Bell and Lloyd 2015), as node ages should only be influenced by the root age's offset if they are part of an unbroken chain of zero-length edges leading to the root, meaning that the divergence dates of more derived nodes are likely to stay the same as the root age offset changes, unlike more root-ward divergences.
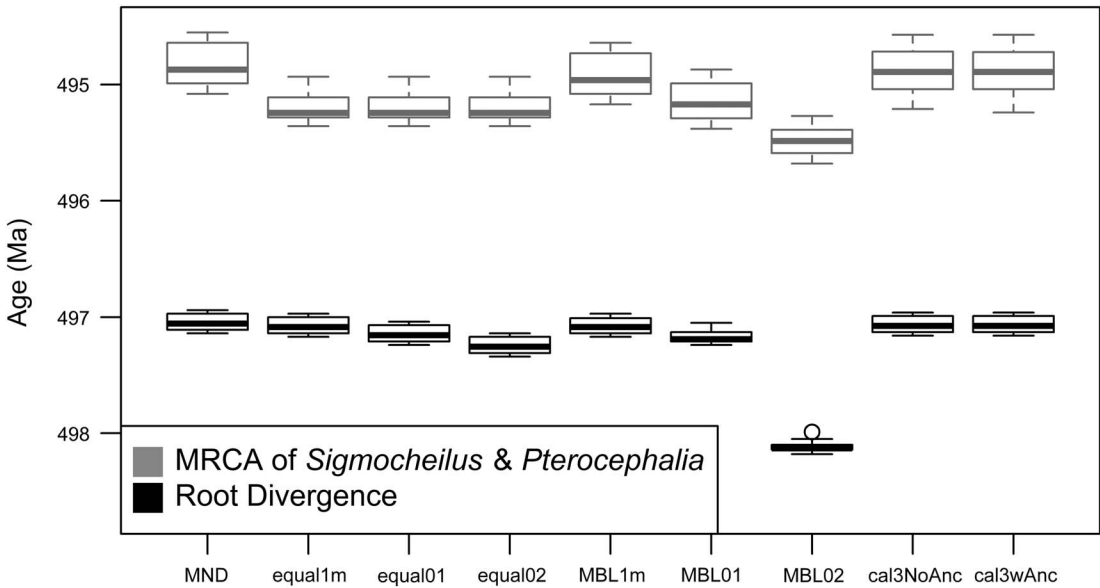
FIGURE 2.   Variation in selected divergence dates under different a posteriori time-scaling approaches. Box plots reflect the distribution of ages for the root node and a nested node that represents the most recent common ancestor (MRCA) of the genera *Sigmocheilus* and *Pterocephalia*, calculated across 100 sets of stratigraphic ranges, reflecting different CONOP solutions. For the time-scaling approaches listed, MND represents minimum-node dating. The "equal1m," "equal01," and "equal02" labels represent the equal branch-sharing approach (Brusatte et al. 2008) with the root age adjusted by 1 m of stratigraphic height (i.e., 0.03 Myr), 0.1 Myr, and 0.2 Myr, respectively. The "MBL1m," "MBL01," and "MBL02" labels represent the minimum branch length approach (Laurin 2004) using minimum lengths of 0.03 Myr, 0.1 Myr, and 0.2 Myr, respectively. See Table 1 for further description of these approaches. The right-most labels, "cal3NoAnc" and "cal3wAnc," represent the *cal3* approach, variously disallowing or allowing for the *cal3* algorithm to infer observed taxa as direct ancestors (Bapst 2013a).

For both of these node age comparisons, *cal3* implies node age distributions that are very similar to the MND approach.

Comparing Support for Early-burst Trait Evolution on Time-Scaled Trees

Hopkins (2013a) reported a "triangle-shaped" association between time and rate of morphological evolution in the pterocephaliids, such that rates were highly variable early on, with the maximum rate decreasing through time. As a result, the original reported correlation was not particularly high (Kendall's tau = −0.24). This value is only slightly more negative than correlation estimates from trees time-scaled the same way (equal, root offset of 1 m) reported herein (Fig. 3A), suggesting that our methodology is consistent with Hopkins's previous protocol (any difference is due to the different range composites used). However, the sign and strength of the correlation is

highly dependent on the time-scaling method (Fig. 3A). Over all samples, MND and *cal3* tend toward zero correlation, equal time-scaling tends to suggest negative correlations, and MBL implies positive correlations. In addition, while the relative difference in correlations mainly stays the same when time-scaled trees instead of a unit-length tree are used for ancestral trait reconstruction (Fig. 3B), all tree samples shift at least marginally toward more positive correlations (i.e., more support for a late-burst scenario), implying that the new ancestral reconstructions imply more trait change on later rather than earlier branches. The largest shift in correlation coefficient occurs with minimum-dated trees: this likely reflects the effect that zero-length time-scaled edges are having on the ancestral trait reconstruction, as they are only removed postreconstruction.

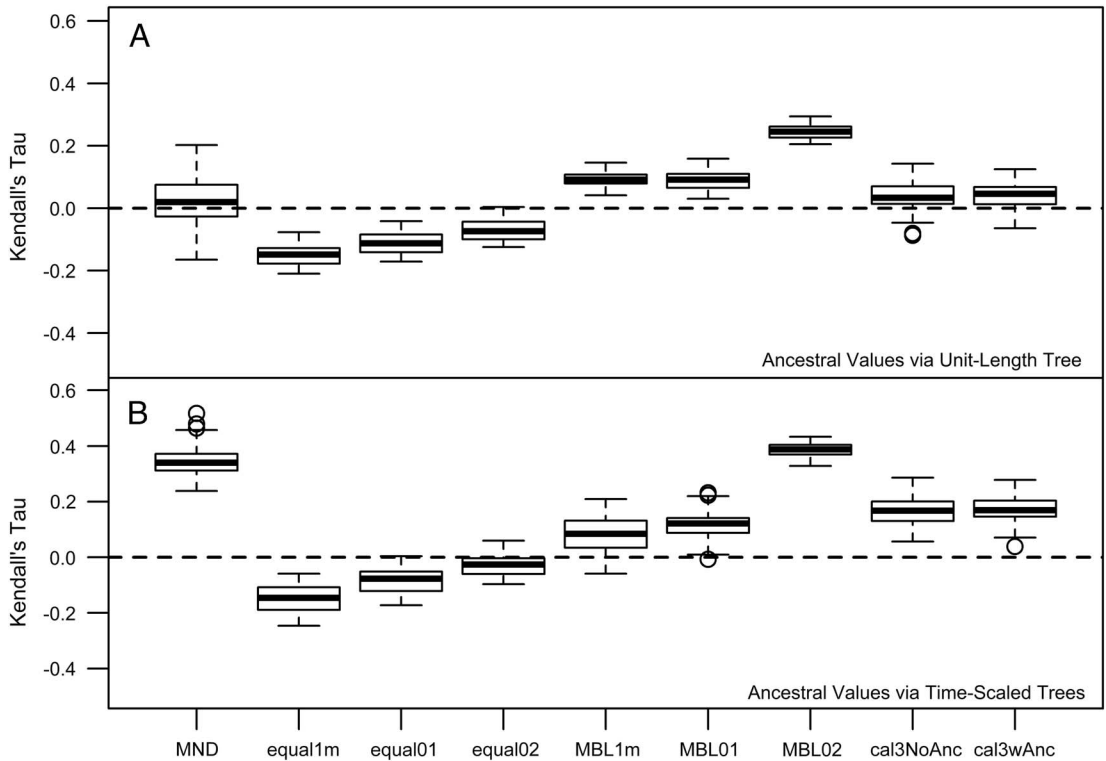Model-fitting analyses also show weak to moderate support for either early-burst or

FIGURE 3. Kendall's tau correlations of the rate ancestor–descendant change (across all PCs) versus node age. Node age was negated, such that positive correlations represent rates increasing with time (late burst) and negative correlations reflect rates decreasing with time (early burst). There were some differences among estimates depending on whether ancestral trait reconstruction was done on a phylogeny in which all edge lengths were scaled to 1 (A) or on the time-scaled tree used for other components of the rate calculation (B). See the caption for Fig. 2 for information on the respective time-scaling approaches.

late-burst scenarios, but some dating methods gave contradictory inferences relative to those from the ancestor–descendant contrast analyses (Fig. 4). In particular, the MND and cal3 approaches suggested late-burst dynamics using ancestor–descendant contrasts on dated trees, but early-burst dynamics when model-fitting analyses were applied. While MND trees imply a wide degree of uncertainty for either scenario, other methods imply between one to two doublings and/or halvings of the evolutionary rate over 5 Myr, with that envelope of uncertainty being largest under *cal3*, which weakly supports an early-burst scenario.

Inferring Ancestor–Descendant Relationships with *cal3*

Based on tree topology, stratigraphic distribution, and optimized character state

transformations, Hopkins (2011a) named nine potential ancestor–descendant pairs in the pterocephaliids, including six budding clado-genetic pairs (in which a morphologically distinguishable daughter species appears via branching from some persistent ancestral species) and three anagenetic pairs (in which one morphologically defined species evolves, without branching, into another morpho-species). We recovered 16 potential ancestor–descendant pairs across 100 trees generated using *cal3* and last appearance dates, with each tree containing 8 to 14 of these pairs. Thirteen pairs were found across more than 60% of the *cal3* trees (Fig. 5), which includes all nine a priori pairs listed by Hopkins (2011b). However, anagenesis was rarely inferred for any species pair, even for those previously supposed to represent anagenesis. Almost all inferred ancestors were placed as ancestor to
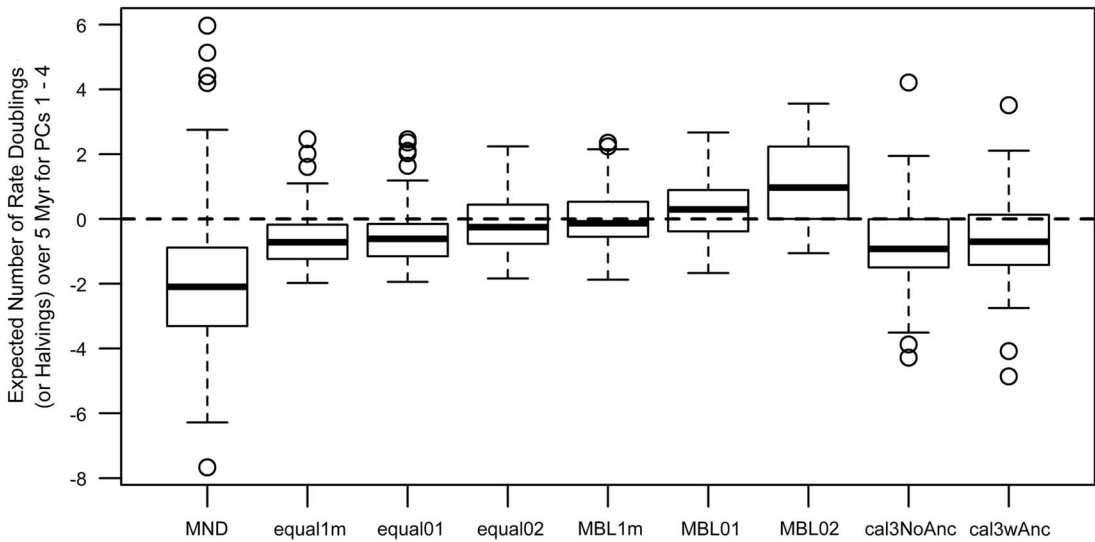
FIGURE 4. Estimates of time-dependent change in the rates of trait evolution (across PCs 1–4) using multivariate model-fitting analyses from R package mvMORPH (Clavel et al. 2015). The summary statistics shown are based on estimated model parameters (Slater and Pennell 2014) and represent, if positive, the expected number of doublings of rate expected over 5 million years, indicating that the rate of trait evolution has increased over time (late burst). If negative, the values represent the expected number of halvings of rate over 5 Myr, indicating that the rate of trait evolution has decreased over time (early burst). See the caption for Fig. 2 for information on the respective time-scaling approaches.

a single descendant taxon, reflecting that *cal3* is very constrained by the cladogram used, which lacks polytomies, and thus any given taxon can only be ancestral to its respective sister taxon. The only variation that occurs is when there is variation in the order of appearance of taxa across CONOP solutions. For example, *Dicanthopyge quadrata* is listed as an ancestor to both *D. reductus* and *D. convergens*, depending on which taxon is assigned an earlier first appearance time in a composite section. Similarly, whether *Strigambitus utahensis* is an ancestor to *St. transversus* or vice versa depends on which is assigned an earlier first appearance time.

Discussion

The *cal3* time-scaling method requires estimates of diversification and sampling that may, at first glance, seem difficult to estimate. The structure of paleontological data sets can vary greatly, and this makes it difficult to prescribe any simple cookbook procedure for obtaining such estimates. Here, we have obtained meaningful estimates from a data set with qualities that made estimation less than

straightforward by taking multiple analytical approaches (Fig. 1) and through careful comparison to previous published estimates. By applying *cal3* and other time-scaling methods to this data set, we find that choice of time-scaling method can impact the estimates of divergence dates (Fig. 2) and downstream macroevolutionary analyses (Figs. 3, 4). The *cal3* approach also allows us to consider the potential for sampled ancestors in our data set, which is not offered by any other APT approach (Fig. 5). Furthermore, as the results of analyses from multiple MPT and branch-sharing analyses did not always include the results of the *cal3* analyses, it may be inadequate to test for sensitivity to dating error by using various arbitrary APT approaches with a range of constants (e.g., Betancur-R et al. 2015; Bates et al. 2016).

Furthermore, the varying results of the analyses we applied to test for early bursts of rapid trait evolution in the pterocephaliids were not due only to varying time-scaling methods, but also to the specific approach used to analyze trait evolution. In particular, trees dated with *cal3* predicted a weak late-burst pattern using the ancestor-reconstruction
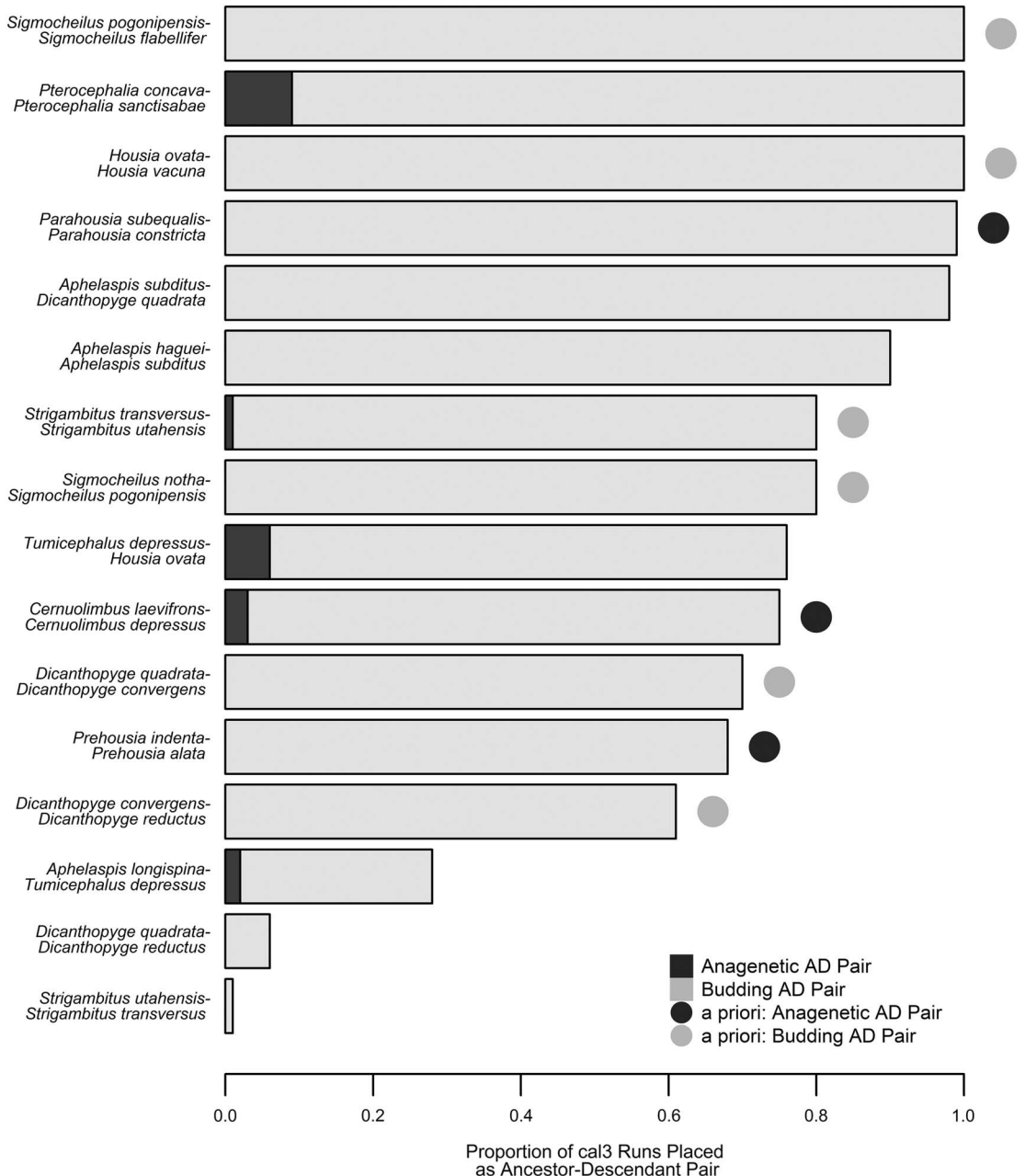
FIGURE 5. Comparison of inferred ancestor–descendant pairs via *cal3* with a priori expectations. The proportion of *cal3* time-scaled trees that included a particular ancestor–descendant pair is listed as a stacked bar plot, with the evolutionary mode (anagenesis or budding cladogenesis; Foote 1996) indicated by color (dark gray or light gray, respectively). Pairs expected based on previous systematic work (Hopkins 2011b) are denoted by a circle adjacent to the respective bar, with identical color coding for each evolutionary model.

approach, but a weak early-burst pattern using the model-fitting approach. Ancestral reconstructions of continuous traits are often avoided by evolutionary biologists because of their wide confidence intervals, suggesting a high degree of uncertainty (Schluter et al. 1997; Losos 2011). Although it has been shown that fossils can improve the accuracy of

reconstructed trait values (Finarelli and Flynn 2006), there is as yet no evidence that fossils improve precision to such a degree that they become usable for testing macroevolutionary patterns (contra Alroy 2000), despite their wide use in paleobiology (Butler and Goswami 2008; Zanno and Mackovicky 2013; Hopkins 2016). In this case, there is no way to determine which analytical approach is more accurate, and regardless of the evolutionary interpretation of directionality, the difference in effect size from time-constant rates is relatively small. We recommend that paleontologists using phylogenetic comparative methods to test macroevolutionary hypotheses should always explore multiple approaches.

The *cal3* approach can infer ancestor–descendant pairs and categorizes these between anagenetic and budding cladogenetic relationships, and it may help unlock the extent to which branching events are integral to how morphotaxa "speciate" (morphologically differentiate) from one another (Wagner and Erwin 1995; Ezard et al. 2012; Bapst 2013b; Pennell et al. 2014). However, the *cal3* determination is independent of character data and is based solely on stratigraphic ranges and tree topology. Tip-dating methods can also place ancestors, but only as sampled ancestral populations assigned to a specific point in time (Gavryushkina et al. 2014; Heath et al. 2014), and thus cannot distinguish modes, while stratocladistic methods can infer mode but are not based on an explicit probabilistic model (Fisher 2008). The ancestor–descendant pairs inferred by *cal3* and presumed a priori match fairly well, suggesting that *cal3* ancestor inference is congruent with qualitative assessments of ancestral relationships. The only seeming discordance is the low observation of anagenetic ancestors, given that two of the anagenetic pairs proposed by Hopkins (2011b) have simultaneous last appearance dates of the ancestor and first appearance dates of the descendant in the majority of our 100 range composites. However, distinguishing between inferred anagenetic and cladogenetic pairs does not directly equate to the true proportion of anagenesis to cladogenesis. Some "apparent" anagenetic relationships are likely cladogenetic. Anagenesis might be more

apparent than real when a taxon at its last appearance time is inferred to be in a direct line of ancestry to one or more sampled taxa, as this pattern would also be observed where branching (i.e., budding or bifurcating cladogenesis; Foote 1996) occurs after that last appearance. Any case in which sister taxon ranges do not overlap could potentially be a case of apparent anagenesis (but see Ezard et al. 2012 for a contrarian view), with no further grounds to confirm or reject the pattern of differentiation between a particular pair of taxa. This effect is asymmetrical: apparent budding cladogenesis supports only budding cladogenesis, not anagenesis or bifurcation. Thus, there should be an expectation that a number of budding or bifurcating cladogenetic events are mistakenly categorized as anagenesis in any incompletely sampled fossil record. Less than 2% of the ancestor–descendant relationships were categorized as apparent anagenesis, which might be consistent with morphospecies originating only via budding cladogenesis, not true anagenesis or bifurcating cladogenesis. Such a scenario would agree with previous estimates (Wagner and Erwin 1995). As current tip-dating approaches only treat taxon units as instantaneous populations, they are not able to differentiate between budding and anagenetic relationships, and thus this sort of analysis is currently possible only with *cal3*.

Summary

Obtaining accurate, dated phylogenies for fossil taxa is a major objective of modern paleontological systematics. Development of new methods has accelerated in recent years and will probably continue to accelerate for some time into the future. The invention of new probabilistic APT methods, like *cal3* (Bapst 2013a), and sampled-ancestor tip-dating (Gavryushkina et al. 2014; Zhang et al. 2016) allows workers to avoid arbitrary time-scaling approaches. Eventually, APT approaches may be replaced by Bayesian tip-dating with fixed topologies and no character data, as already seen in some vertebrate paleobiology studies (e.g., Close et al. 2015; Fischer et al. 2016), in which taxa are known from a small

number of dated collections. Despite these methodological advances, the simpler APT approaches remain in common use, even though it is clear from both previous simulations (Bapst 2014) and empirical case studies (such as this paper) that those approaches provide results that differ from statistically valid time-scaling approaches. We show here that it is possible to get useful estimates of sampling, branching, and extinction rates for *cal3* and that it is worth the effort. A dated phylogeny of fossil taxa is a complex product that can have considerable influence on the macroevolutionary inferences made from it. Dated phylogenies of fossil taxa also offer immediate opportunities to compare the estimated divergence dates with the timing of stratigraphic appearances and paleoenvironmental changes and the potential to explore ancestor–descendant relationships. The steps taken to date fossil phylogenies warrant more attention in the literature, and empirical studies whose main aim is to produce dated phylogenies should be viewed as important and worthy stand-alone contributions to paleontological systematics.

Acknowledgments

Literature Cited

Alba, D. M., J. Agustí, and S. Moyà-Solà. 2001. Completeness of the mammalian fossil record in the Iberian Neogene. Paleobiology 27:79–83.

Alroy, J. 2000. Understanding the dynamics of trends within evolving lineages. Paleobiology 26:319–329.

Bapst, D. W. 2012. paleotree: an R package for paleontological and phylogenetic analyses of evolution. Methods in Ecology and Evolution 3:803–807.

——. 2013a. A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. Methods in Ecology and Evolution 4:724–733.

——. 2013b. When can clades be potentially resolved with morphology? PLoS ONE 8:e62312.

——. 2014. Assessing the effect of time-scaling methods on phylogeny-based analyses in the fossil record. Paleobiology 40:331–351.

Bapst, D. W., A. M. Wright, N. J. Matzke, and G. T. Lloyd. 2016. Topology, divergence dates, and macroevolutionary inferences vary between different tip-dating approaches applied to fossil theropods (Dinosauria). Biology Letters 12(7): 20160237. doi: 10.1098/rsbl.2016.0237.

Bates, K. T., P. D. Mannion, P. L. Falkingham, S. L. Brusatte, J. R. Hutchinson, A. Otero, W. I. Sellers, C. Sullivan, K. A. Stevens, and V. Allen. 2016. Temporal and phylogenetic evolution of the sauropod dinosaur body plan. Open Science 3(3): 150636. doi: 10.1098/rsos.150636.

Bell, M. A., and G. T. Lloyd. 2015. strap: an R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence. Palaeontology 58:379–389.

Betancur-R, R., G. Ortí, and R. A. Pyron. 2015. Fossil-based comparative analyses reveal ancient marine ancestry erased by extinction in ray-finned fishes. Ecology Letters 18:441–450.

Blomberg, S. P., T. J. Garland, and A. R. Ives. 2003. Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. Evolution 57:717–745.

Bokma, F., M. Godinot, O. Maridet, S. Ladevèze, L. Costeur, F. Solé, E. Gheerbrant, S. Peigné, F. Jacques, and M. Laurin. 2016. Testing for Depéret's rule (body size increase) in mammals using combined extinct and extant data. Systematic Biology 65:98–108.

Brocklehurst, N. 2016. Rates and modes of body size evolution in early carnivores and herbivores: a case study from Captorhinidae. PeerJ 4:e1555.

Brusatte, S. L., M. J. Benton, M. Ruta, and G. T. Lloyd. 2008. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. Science 321:1485–1488.

Butler, R. J., and A. Goswami. 2008. Body size evolution in Mesozoic birds: little evidence for Cope's rule. Journal of Evolutionary Biology 21:1673–1682.

Clavel, J., G. Escarguel, and G. Merceron. 2015. mvmorph: an R package for fitting multivariate evolutionary models to morphometric data. Methods in Ecology and Evolution 6:1311–1319.

Close, Roger A., M. Friedman, Graeme T. Lloyd, and Roger B. J. Benson. 2015. Evidence for a mid-Jurassic adaptive radiation in mammals. Current Biology 25:2137–2142.

Cody, R. D., R. H. Levy, D. M. Harwood, and P. M. Sadler. 2008. Thinking outside the zone: high-resolution quantitative diatom biochronology for the Antarctic Neogene. Palaeogeography, Palaeoclimatology, Palaeoecology 260:92–121.

Eiting, T., and G. Gunnell. 2009. Global completeness of the bat fossil record. Journal of Mammalian Evolution 16:151–173.

Etienne, R. S., and J. Rosindell. 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. Systematic Biology 61:204–213.

Ezard, T. H. G., T. Aze, P. N. Pearson, and A. Purvis. 2011. Interplay between changing climate and species' ecology drives macroevolutionary dynamics. Science 332:349–351.

Ezard, T. H. G., P. N. Pearson, T. Aze, and A. Purvis. 2012. The meaning of birth and death (in macroevolutionary birth–death models). Biology Letters 8:139–142.

Finarelli, J. A., and J. J. Flynn. 2006. Ancestral state reconstruction of body size in the Caniformia (Carnivora, Mammalia): the effects of incorporating data from the fossil record. Systematic Biology 55:301–313.

Fisher, D. C. 2008. Stratocladistics: integrating temporal data and character data in phylogenetic inference. Annual Review of Ecology, Evolution, and Systematics 39:365–385.

Fischer, V., N. Bardet, R. B. J. Benson, M. S. Arkhangelsky, and M. Friedman. 2016. Extinction of fish-shaped marine reptiles associated with reduced evolutionary rates and global environmental volatility. Nature Communications 7:10825. doi: 10.1038/ncomms10825.

Foote, M. 1996. On the probability of ancestors in the fossil record. Paleobiology 22:141–151.

——. 1997. Estimating taxonomic durations and preservation probability. Paleobiology 23:278–300.

——. 2000. Origination and extinction components of taxonomic diversity: general problems. Pp. 74–102 in D. H. Erwin, and S. L. Wing, eds. Deep time: paleobiology's perspective. Paleontological Society, Lawrence, Kans.

——. 2001. Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. Paleobiology 27:602–630.

Foote, M., and D. M. Raup. 1996. Fossil preservation and the stratigraphic ranges of taxa. Paleobiology 22:121–140.

Foote, M., and J. J. Sepkoski. 1999. Absolute measures of the completeness of the fossil record. Nature 398:415–417.

Foote, M., J. P. Hunter, C. M. Janis, and J. J. Sepkoski, Jr. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. Science 283:1310–1314.

Friedman, M., and M. D. Brazeau. 2011. Sequences, stratigraphy, and scenarios: what can we say about the fossil record of the earliest tetrapods? Proceedings of the Royal Society of London B 278:432–439.

Gavryushkina, A., D. Welch, T. Stadler, and A. J. Drummond. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. PLoS Computational Biology 10:e1003919.

Gradstein, F. M., G. Ogg, and M. Schmitz. 2012. The Geologic Time Scale 2012. Elsevier, Oxford.

Halliday, T. J. D., and A. Goswami. 2016a. Eutherian morphological disparity across the end-Cretaceous mass extinction. Biological Journal of the Linnean Society 118:152–168.

——. 2016b. The impact of phylogenetic dating method on interpreting trait evolution: a case study of Cretaceous–Palaeogene eutherian body-size evolution. Biology Letters 12(8): 20160051. doi: 10.1098/rsbl.2016.0051.

Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences USA 111:E2957–E2966.

Holland, S. M. 2003. Confidence limits on fossil ranges that account for facies changes. Paleobiology 29:468–479.

——. 2016. The non-uniformity of fossil preservation. Philosophical Transactions of the Royal Society of London B 371(1699): 20150130. doi: 10.1098/rstb.2015.0130.

Hopkins, M. J. 2011a. How species longevity, intraspecific morphological variation, and geographic range size are related: a comparison using Late Cambrian trilobites. Evolution 65: 3253–3273.

——. 2011b. Species-level phylogenetic analysis of pterocephaliids (Trilobita, Cambrian) from the Great Basin, western USA. Journal of Paleontology 85:1128–1153.

——. 2013a. Data from: Decoupling of taxonomic diversity and morphological disparity during decline of the Cambrian trilobite family Pterocephaliidae. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.f5667.

——. 2013b. Decoupling of taxonomic diversity and morphological disparity during decline of the Cambrian trilobite family Pterocephaliidae. Journal of Evolutionary Biology 26: 1665–1676.

——. 2016. Magnitude versus direction of change and the contribution of macroevolutionary trends to morphological disparity. Biological Journal of the Linnean Society 118:116–130.

Hopkins, M. J., and A. B. Smith. 2015. Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution. Proceedings of the National Academy of Sciences USA 112:3758–3763.

Hunt, G., and M. T. Carrano. 2010. Models and methods for analyzing phenotypic evolution in lineages and clades. Pp. 245–269 in J. Alroy, and G. Hunt, eds. Short Course on Quantitative Methods in Paleobiology. Paleontological Society, New Haven, Conn.

Kendall, D. G. 1948. On the generalized "birth-and-death" process. Annals of Mathematical Statistics 19(1): 1–15.

Laurin, M. 2004. The evolution of body size, Cope's rule and the origin of amniotes. Systematic Biology 53:594–622.

Lee, M. S. Y., A. Cau, D. Naish, and G. J. Dyke. 2014. Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. Science 345:562–566.

Losos, J. B. 2011. Seeing the forest for the trees: the limitations of phylogenies in comparative biology. American Naturalist 177:709–727.

Ludvigsen, R., and S. R. Westrop. 1985. Three new Upper Cambrian stages for North America. Geology 13:139–143.

Nee, S. 2006. Birth–death models in macroevolution. Annual Review of Ecology, Evolution, and Systematics 37:1–17.

Nee, S., A. O. Mooers, and P. H. Harvey. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proceedings of the National Academy of Sciences USA 89:8322–8326.

Norell, M. A. 1992. Taxic origin and temporal diversity: the effect of phylogeny. Pp. 89–118 in M. J. Novacek, and Q. D. Wheeler, eds. Extinction and phylogeny. Columbia University Press, New York.

Palmer, A. R. 1962. Glyptagnostus and associated trilobites in the United States. U.S. Geological Survey Professional Paper 374-F.

——. 1965. Trilobites of the late Cambrian Pterocephaliid biomere in the Great Basin, United States. U.S. Geological Survey Professional Paper 493.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Pennell, M. W., L. J. Harmon, and J. C. Uyeda. 2014. Is there room for punctuated equilibrium in macroevolution? Trends in Ecology and Evolution 29:23–32.

Peters, S. E., and M. Foote. 2001. Biodiversity in the Phanerozoic: a reinterpretation. Paleobiology 27:583–601.

Puttick, M. N., G. H. Thomas, and M. J. Benton. 2016. Dating Placentalia: morphological clocks fail to close the molecular fossil gap. Evolution 70:873–886.

Pyron, R. A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Systematic Biology 60:466–481.

R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raup, D. M. 1985. Mathematical models of cladogenesis. Paleobiology 11:42–52.

Raup, D. M., S. J. Gould, T. J. M. Schopf, and D. S. Simberloff. 1973. Stochastic models of phylogeny and the evolution of diversity. Journal of Geology 81:525–542.

Ronquist, F., S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. L. Murray, and A. P. Rasnitsyn. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology 61:973–999.

Ruta, M., P. J. Wagner, and M. I. Coates. 2006. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. Proceedings of the Royal Society of London B 273:2107–2111.

Sadler, P. M. 2007. Constrained optimization approaches to the paleobiologic correlation and seriation problem. Part II: a reference manual to the CONOP program family. http://www.springer.com/us/book/9781402014437.

Sadler, P. M., W. G. Kemple, and M. A. Kooser. 2003. Contents of the compact disc: CONOP9 programs for solving the stratigraphic correlation and seriation problems as constrained optimization. In P. J. Harries, ed. Approaches in high-resolution stratigraphic paleontology. Topics in Geobiology 21. Kluwer Academic, Dordrecht, Netherlands.

Schluter, D., T. Price, A. O. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711.

Sepkoski, J. J. 1998. Rates of speciation in the fossil record. Philosophical Transactions of the Royal Society of London B 353:315–326.

Silvestro, D., J. Schnitzler, L. H. Liow, A. Antonelli, and N. Salamin. 2014. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. Systematic Biology 63: 349–367.

Slater, G. J., and M. W. Pennell. 2014. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. Systematic Biology 63:293–308.

Smith, A. B. 1994. Systematics and the fossil record: documenting evolutionary patterns. Blackwell Scientific, Oxford.

Solow, A. R., and W. Smith. 1997. On fossil preservation and the stratigraphic ranges of taxa. Paleobiology 23:271–277.

Soul, L. C., and M. Friedman. 2015. Taxonomy and phylogeny can yield comparable results in comparative paleontological analyses. Systematic Biology 64:608–620.

Stadler, T. 2010. Sampling-through-time in birth-death trees. Journal of Theoretical Biology 267:396–404.

Stadler, T., and Z. Yang. 2013. Dating phylogenies with sequentially sampled tips. Systematic Biology 62:674–688.

Stanley, S. M. 1979. Macroevolution: patterns and process. Freeman, San Francisco.

Starrfelt, J., and L. H. Liow. 2016. How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model. Philosophical Transactions of the Royal Society of London B 371(1691): 20150219. doi: 10.1098/rstb.2015.0219.

Strauss, D. J., and P. M. Sadler. 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology 21:411–427.

Tomiya, S. 2013. Body size and extinction risk in terrestrial mammals above the species level. American Naturalist 182:E196–E214.

Vandenberg, A. H. M. 2003. Discussion of "Gisbornian (Caradoc) graptolites from New South Wales, Australia: systematics, biostratigraphy and evolution" by B. Rickards, L. Sherwin and P. Williamson. Geological Journal 38:175–179.

Wagner, P. J., and D. H. Erwin. 1995. Phylogenetic patterns as tests of speciation models. Pp. 87–122 in D. H. Erwin, and R. L. Anstey, eds. New approaches to speciation in the fossil record. Columbia University Press, New York.

Wagner, P. J., and J. D. Marcot. 2010. Probabilistic phylogenetic inference in the fossil record: current and future applications. Pp. 189–211. in J. Alroy, and G. Hunt, eds. Short Course on Quantitative Methods in Paleobiology. Paleontological Society, New Haven, Conn.

——. 2013. Modelling distributions of fossil sampling rates over time, space and taxa: assessment and implications for macroevolutionary studies. Methods in Ecology and Evolution 4:703–713.

Wright, A. M., G. T. Lloyd, and D. M. Hillis. 2016. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. Systematic Biology 65:602–611.

Zanno, L. E., and P. J. Makovicky. 2013. No evidence for directional evolution of body mass in herbivorous theropod dinosaurs. Proceedings of the Royal Society of London B 280(1751): 20122526. doi: 10.1098/rspb.2012.2526.

Zhang, C., T. Stadler, S. Klopfstein, T. A. Heath, and F. Ronquist. 2016. Total-evidence dating under the fossilized birth–death process. Systematic Biology 65:228–249.