Syst. Biol. 00(0):1–17, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of the Society of Systematic Biologists. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site-for further information please contact journals.permissions@oup.com
https://doi.org/10.1093/sysbio/syaf078
Advance Access publication 0, 2025

Incorporating Continuous Characters in Joint Estimation of Dicynodont Phylogeny

Brenen M. Wynd[1], Basanta Khakurel[1,2,3], Christian F. Kammerer[4,5], Peter J. Wagner[6], and April M. Wright[1,*]

[1]*Department of Biological Sciences, Southeastern Louisiana University, Hammond, LA, USA*
[2]*GeoBio-Center, Ludwig-Maximilians-Universität München, Richard-Wagner-Str. 10, 80333 Munich, , Germany*
[3]*Department of Earth and Environmental Sciences, Paleontology & Geobiology, Ludwig-Maximilians-Universität München, Richard-Wagner-Str. 10, 80333 Munich, Germany*
[4]*North Carolina Museum of Natural Sciences, Raleigh, NC, USA*
[5]*Department of Biological Sciences, North Carolina State University, Raleigh, NC, USA*
[6]*Department of Earth & Atmospheric Sciences, and School of Biological Sciences, University of Nebraska, Lincoln, NE, USA*
[*]*Correspondence to be sent to: Department of Biological Sciences, Southeastern Louisiana University, Hammond, LA, USA; E-mail: april.wright@southeastern.edu*

Received 25 February 2025; reviews returned 3 October 2025; accepted 14 October 2025
Associate Editor: Joëlle Barido-Sottani

Abstract.—Continuous characters have received comparatively little attention in Bayesian phylogenetic estimation. This is predominantly because they cannot be modeled by a standard phylogenetic Q-matrix approach due to their non-discrete nature. In this paper, we explore the use of continuous traits under two Brownian motion models to estimate a phylogenetic tree for Dicynodontia, a well-studied group of early synapsids (stem mammals) in which both discrete and continuous characters have been extensively used in parsimony-based tree reconstruction. We examine the differences in phylogenetic signal between a continuous trait partition, a discrete trait partition, and a joint analysis with both types of characters. We find that continuous and discrete traits contribute substantially different signal to the analysis, even when other parts of the model (clock and tree) are held constant. Tree topologies resulting from the new analyses differ strongly from the established phylogeny for dicynodonts, highlighting continued difficulty in incorporating truly continuous data in a Bayesian phylogenetic framework. [Bayesian phylogenetics; continuous traits; discrete traits; morphological data; phylogenetic methods; RevBayes.]

Phylogenetics strives to use organismal data that reflect common inheritance and are directly comparable across taxa in the sample set. In working with extant taxa, molecular information (DNA, RNA, and amino acid sequences) can offer objective measures that can be compared across species. A single nucleotide base is generally assumed to have the same chemical structure and properties as all other nucleotides of the same type in an alignment or across species. The same is not true for anatomical data (Lewis 2001; Huelsenbeck and Rannala 2003; Scotland et al. 2003; Brazeau 2011; Klopfstein et al. 2015; Wright et al. 2016; Simões et al. 2023; Khakurel et al. 2024), for which structure and expression is often unique not only for individual taxa, but is also filtered through human perception and interpretation and thus varies between individual researchers. Morphological character states almost never reflect truly identical structures everywhere they appear in a character matrix (Lewis 2001), and different authors may code similar structures differently (Jardine 1969). Even among presence/absence characters, though "absent" is (on one level) objectively comparable, "present" often represents different and not necessarily equivalent additions to the organisms' morphology. These types of characters are ubiquitous: across sampled morphological data sets, roughly 70% of characters are binary in this manner (Barido-Sottani et al. 2020). Multistate discrete characters represent their own difficulties in terms of repeatability. State spaces may be misspecified, leading to incorrect homology statements and model parameterization (Hoyal Cuthill 2015; Khakurel et al. 2024; Mulvey et al. 2024).

Historically, discrete morphological characters are heavily represented in fields that lack ready access to DNA (e.g., paleontology) because of their utility in parsimony-based cladistics (Wright and Wynd 2024), a tree-estimation criterion that seeks to generate the most likely tree by finding the fewest number of transitions possible among characters (i.e., Occam's Razor). Parsimony has been the dominant phylogenetic tool for paleontologists. With the growing usage of Bayesian methods in morphological phylogeny estimation, Bayesian estimation methods have been evaluated for discrete characters (Lewis 2001; Wright and Hillis 2014; Klopfstein et al. 2015, 2019; O'Reilly et al. 2016; Wright et al. 2016). Because of their ready availability from prior discrete data sets intended for parsimony-based analyses, authors have been able to evaluate these methods using both simulated and empirical data sets. Though the methods themselves have been evaluated, there are still considerable questions surrounding how morphological data should be coded for Bayesian phylogenetic analysis. The fundamental units (such as "0"

or "1") are not necessarily comparable across a data matrix (Pogue and Mickevich 1990; Lewis 2001; Nylander et al. 2004). For example, a muscle scar having a relatively larger attachment area may be considered, *a priori*, to be faster evolving or "easier" to evolve than the introduction of a brand new muscle scar, but in a parsimony framework, they would both be scored as "present" (1) and would be analytically indistinguishable from one another. How discretizing data in matrix assembly, and what information is lost in doing so, affects overall estimation is still an open question.

Discrete morphological characters often seek to represent complex and difficult-to-measure traits as binary or multistate characters, with numeric states corresponding to a more precise character definition. The alternative to discretization is quantifying traits via direct measurement of dimensions or proportions into continuous variables. Ratios are another type of continuous trait, and tend to be favored in some circumstances as raw measurements are more likely to be influenced by ontogeny when a sample includes differently sized individuals (Bardin et al. 2017; Griffin et al. 2021; Napoli 2024). Direct measurements of anatomical features, such as lengths or angulations of structures or ratios of one structure to another, have the potential to be more objective than qualitative character state assignments, although worker choice in what features to measure remains a source of some bias. Continuous trait use in phylogenetics has been controversial over the years, with some authors arguing that continuous traits are too noisy for phylogenetics (Chappill 1989), whereas others argue that they are the "Holy Grail" of phylogenetic characters (Thiele 1993). In recent years, there has been a renewed interest in continuous characters from biology and anthropology (Parins-Fukuchi 2018; Matzig et al. 2024; Zhang et al. 2024). Continuous variables are difficult to model in phylogenetics, as the degree of change over a single unit of time can be infinitesimal and thus are ill-equipped for the transition matrices that form the basis of both cladistic and likelihood-based frameworks. Perhaps as a result of both issues, continuous characters are comparatively rare in the phylogenetic literature. For example, in Graeme Lloyd's repository of character matrices (Wright et al. 2016), a commonly used database of morphological data sets, only 10 out of 573 matrices have any continuous traits. The others rely solely on discrete characters.

Many of the objections to and difficulties in using continuous anatomical measurements in phylogenetic analyses are predicated upon using minimum-steps parsimony methods to infer phylogeny, and thus might not pertain to likelihood and Bayesian analyses. Although continuous traits constitute a minority of data sets, parsimony-based cladistics is able to incorporate continuous variables via the additive method (Farris 1970), which samples sister taxon relationships from a given character and seeks to minimize the difference between nodes moreso than tips. Conceptually, the additive method can be thought of as an inversion of Felsenstein's independent contrasts (Felsenstein 1985a). In an additive model, trait space cannot be explored by jumping over intermediate states (Goloboff et al. 2006). The additive model for continuous characters, while not explicitly modeled as such, essentially follows a single rate Brownian motion (BM) model (i.e., random walk) for evolution (Felsenstein 1985b), where the intermediate states are dependent on the terminal states, such that the model retains memory (Goloboff et al. 2006). However, instead of the amount of change per unit time being fixed to a specific rate (σ) that applies to branch lengths as in a BM model, it was instead drawn from a distribution formed by the range of tip states for the descendants of the node in question (Farris 1970). Although the philosophies behind BM and additive character evolution models are quite similar, they are computationally distinct, as BM is reliant on dated trees with informative branch lengths, and additive models only estimate intermediate states at nodes and are generally applied to undated trees (Farris 1970; Hunt 2007). This similarity thus suggests that modeling continuous characters under BM would likely be more suitable in a Bayesian framework, wherein the estimation of branch lengths is essential to the estimation of the tree.

The evolution of continuous variables under comparative methods has primarily been modeled under BM (i.e., random-walk; Felsenstein 1973, 1985a; Gingerich 1993) and Ornstein–Uhlenbeck (OU) (i.e., directional evolution; Hansen 1997; Butler and King 2004; Beaulieu et al. 2012), with many variants existing to relax or constrain model assumptions. These models allow for changes to accumulate continuously along branches, and because BM character evolution models conceptually mirror the additive method, character evolution models should be of similar utility in Bayesian inference as in parsimony. However, these analyses are typically post hoc, and assume a tree known without error. Few have explored how these methods might behave in Bayesian or likelihood estimation of the tree using either empirical or simulation study. Parins-Fukuchi (2018) used a BM (Felsenstein 1973) character evolution model in `RevBayes` (Höhna et al. 2016) to estimate trees from simulated morphological characters, demonstrating their utility in this type of analysis. Further work by Zhang et al. (2024) expanded upon this work with more general frameworks for adding continuous characters to trees. BM generally assumes that a trait will drift from its origin value with rate σ. This is a lightweight model, adding only one parameter, though it can be parameterized such that each character can have its own σ or σ can vary across branches on the tree.

Here, we explore the utility of using character evolution models of BM in estimating tip-dated trees in a Bayesian framework under a fossilized birth–death (FBD) model (Fig. 1). Our dated analysis can be readily separated into its component models: the substitution, clock, and tree models (Warnock and Wright 2020).

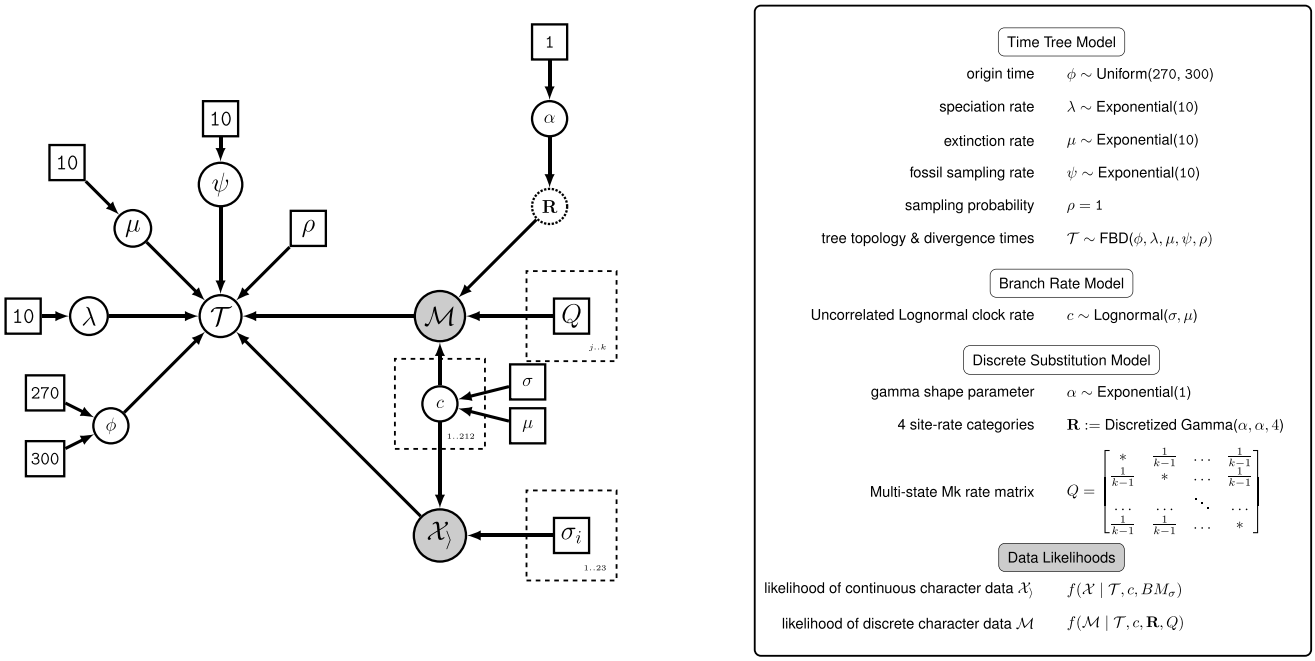


FIGURE 1. A graphical model displaying the components of the full joint phylogenetic model. An explanation of the model components is displayed at right. Adapted from Barido-Sottani et al. (2020).

To jointly estimate the topology and node ages using the FBD, researchers generally assume a tripartite model of evolution: the first that describes the accumulation of character data, the second that describes the evolutionary rates, and the third that describes the speciation, extinction, and sampling events that generate the tree. The benefit of this type of joint model is that the researchers can treat each of the components as a discrete inferential module and also provide the flexibility to combine models according to the data in use (though see May and Rothfels 2023 for complications). This type of approach also allows for inference based on multiple types of data, including molecular or morphological character information. The model components are derived directly from the data, which can be multifarious and for our purposes here will be focused on morphological characters and fossil age information. In our case, this tripartite framework is a particular benefit because we can hold the clock model and tree model constant, while varying the input data and character models to investigate how combinations of data and model assumptions impact the resultant topology.

DICYNODONTIA AS A STUDY SYSTEM

For these analyses, we rely on a recent phylogeny of Dicynodontia, a group of early synapsids (the mammalian total group) originating during the Permian Period. Dicynodonts were a diverse clade of herbivores known from the middle Permian through the Late Triassic periods (ca. ~267–201.5 Ma), and can be readily recognized by their beaked jaws, usually without teeth other than the prominent maxillary tusks, efficient palinal feeding apparatus, wide skulls, and robust bodies (Angielczyk et al. 2017; Sulej and Niedźwiedzki 2019). Over the past 30 years, the phylogenetic history of dicynodonts has been systematically untangled, with intensive alpha taxonomic revision that has reduced the overall species richness of the clade, but increased their utility for Permo-Triassic biostratigraphy (Kammerer et al. 2011; Viglietti et al. 2016; Wynd et al. 2017). Even following the extensive synonymization of historical taxa, however, dicynodonts remain one of the most speciose Permo-Triassic tetrapod groups, with over 100 valid species known globally (Fröbisch 2009). The majority of dicynodont species richness occurs in two major radiations, the Permian therochelonians and their descendants, the Triassic kannemeyeriiforms. Permian and Early Triassic dicynodonts have an excellent fossil record, related to the exceptional nature of the Karoo Basin of South Africa preserving near-continuous sedimentation and capture of associated faunas during this interval (Smith et al. 2020). Most Permian dicynodonts are known minimally from complete skulls, many are known from articulated skeletons, and some, such as *Diictodon* and *Lystrosaurus*, have sample sizes that are remarkable for a pre-Cenozoic terrestrial vertebrate (in the thousands of individuals) (Sullivan et al. 2003; Smith et al. 2013).

Dicynodontia represents an ideal test case to evaluate the impact of continuous characters on morphological character matrices, as they are completely extinct (necessitating the use of morphological data in phylogeny

reconstruction), have a well-documented fossil record, and both continuous traits and discrete traits have been collected for this group and used in previous estimation of dicynodont phylogeny (Angielczyk and Kurkin 2003; Kammerer et al. 2011, 2013, 2019). Current dicynodont phylogeny is the cumulative product of untangling over a century of taxonomic overprinting. The first iteration of the dicynodont data set analyzed herein was produced as part of a monograph revising the genus *Dicynodon*, which is known from hundreds of specimens from the late Permian previously attributed to more than 30 different species (Kammerer et al. 2011). Prior to the reassessment of *Dicynodon*, analyses of dicynodont phylogeny largely focused on either Permian or Triassic taxa in isolation, with examples covering both periods either incorporating very few taxa from the non-focal period or generating composite trees from separate analyses (for a deeper discussion on the origins of this tree, see the section "Phylogenetic Analysis" in Kammerer et al. 2011 and the associated citations therein).

All analyses of dicynodont phylogeny from the past decade are based on the underlying data set of Kammerer et al. (2011), which introduced continuous character data to this system. The goal of the Kammerer et al. (2011) data set was to broadly sample dicynodont diversity, providing for the first time nearly comprehensive coverage of the clade throughout its Permian and Triassic history. Making up this data set were a combination of novel characters and characters adapted from previous analyses of dicynodont phylogeny (e.g., Maisch 2002; Damiani et al. 2007), encompassing the wide variety of morphologies observed throughout anomodont evolution. However, due to variation in the quality of the dicynodont fossil record over time, different character types vary in importance across different groups of dicynodonts.

Continuous variables were introduced to the data set to address historical inconsistencies in discretizing numerous features better explained as measurements and ratios than discrete-state characters. When incorporated under an additive model alongside morphological character data in parsimony-based phylogenetics, these characters have been demonstrated to provide additional phylogenetic resolution unavailable using only discrete-state characters (Kammerer et al. 2011). This is especially true for taxa with very incomplete codings for the majority-cranial discrete-state character set. The later Triassic kannemeyeriiforms (particularly the Stahleckeriidae) tend to be known from less complete crania than their Permian predecessors, with greater representation of taxa based on fragmentary postcrania instead of complete skulls or articulated skeletons. Some of these taxa are currently represented only by single damaged skeletal elements (e.g., *Eubrachiosaurus browni*) or extensive but eroded and disarticulated remains (e.g., *Placerias hesternus*) (Fiorillo et al. 2000; Kammerer et al. 2013). Observed differences in postcranial anatomy between these taxa are largely proportional, and continuous character measurements (primarily ratios) have been an important component of character support in this part of the tree. This offers a unique opportunity to explore which group of characters is contributing signal to the estimation process and how conceptually similar models results differ from one another. Not only will this work help establish future models in phylogenetic estimation, but also we have the unique opportunity to question our established methods and whether or not they behave the ways we expect.

Materials and Methods

Character Evolution Models

Discrete characters: Mk model.— Discrete characters are generally modeled using the Mk model in a likelihood-based framework. The Mk model (Lewis 2001) is a Markov model for morphological evolution, serving as a morphological analogue to the Jukes–Cantor model (Jukes and Cantor 1969). Due to the inherent complexity and diversity of morphological evolution across different characters, the assumptions of the Mk model are inherently simplistic. Notably, this model assumes that the rate of change between the character states is equal.

In the Mk model, the k denotes the number of possible states for a given character, allowing the model to be flexible and accommodate varying numbers of character states across different characters. The rate matrix (Q-matrix) for this model can be generalized based on the number of character states, as follows:

$$Q_k = \begin{pmatrix} * & \frac{1}{k-1} & \cdots & \frac{1}{k-1} \\ \frac{1}{k-1} & * & \cdots & \frac{1}{k-1} \\ \cdots & \cdots & \ddots & \cdots \\ \frac{1}{k-1} & \frac{1}{k-1} & \cdots & * \end{pmatrix}. \quad (1)$$

To address the biological oversimplifications inherent in the Mk model, various extensions have been proposed. For instance, Nylander et al. (2004) introduced priors on the stationary frequencies to relax the assumption of equal transition rates, but were not used in this paper, as prior work demonstrates that character model misspecification is largely unproblematic in dated analyses (Klopfstein et al. 2019).

We employed the Mk (equal-rates) model to generate the substitution (Q) matrices for each character. Partitions were made that separate each character based on the number of character states (Khakurel et al. 2024; Mulvey et al. 2024), to establish n Q-matrices, where n is the maximum number of states for any character in the data set. For the Mk model, we established among-character rate variation by using gamma-distributed rate variation with four discrete rate categories (Yang 1994).

Brownian motion.— The BM model (Felsenstein 1973), also known as the random walk model, assumes that trait evolution occurs through random, incremental changes over time, with both the direction and magnitude of each change governed by a stochastic process. This model includes a single parameter, σ, which represents the evolutionary rate and dictates the size of the phenotypic shift in each time step. Under this model, the phenotypic change in a descendant lineage is assumed to follow a normal distribution, with the mean centered on the ancestral state and the variation proportional to the product of σ and branch length. Consequently, traits are expected to exhibit greater changes along longer branches, reflecting the increased time elapsed since divergence from a common ancestor. This model has been used inferentially in prior studies (Parins-Fukuchi 2018; Zhang et al. 2024).

As a model driven purely by drift, the BM model assumes that the trait evolution is stochastic, independent of other factors such as natural selection or genetic constraints (Butler and King 2004). Despite this simplification, the model remains a valuable tool for investigating general patterns and trends in the evolution of continuous traits over extended time scales.

We used two different types of BM model:

- One-rate BM: Under this model, all characters share a σ value drawn from a uniform distribution bounded between 1e−5 and 1e−1. This model will be referred to as *one-rate* in figures and further text.
- Multi-rate BM: Under this model, each character has an independent σ, drawn from the uniform distribution described above. This model will be referred to as *multi-rate* in figures and further text.

Both models assume uncorrelated characters. This will strike some readers as unrealistic, but for a first-pass application, this seems reasonable, and matched to the assumptions of independence we make for discrete characters.

Fossilized birth–death model.— The FBD model was held constant across all data sets and analyses. For the major parameters of the FBD, we used the Paleobiology Database (PBDB) visual tools (Peters and McClennen 2016; Uhen et al. 2023) to obtain estimates for origination and extinction rates for the clade from published specimens. We sampled speciation rates from an exponential prior with a mean of 5.5, based on this output. Turnover rate was pulled from a uniform distribution between 0.90 and 1.05. From turnover rate and speciation rate, we can estimate extinction rate (turnover − speciation rate) as well as diversification rate (speciation − extinction rate). The fossil sampling rate (ψ) was sampled from an exponential distribution with a mean of (1/0.072), which is the average per-million year expectation given the character frequencies. We set the proportional taxon sampling of the youngest time slice (ρ) at 0.231, based on the distribution of finds among synapsids and archosaurs from sediments contemporaneous with the latest therapsids. Origin time, the time of divergence for the last universal common ancestor for the given topology, was sampled between 270 and 300 Ma. A complete model can be viewed in Figure 1.

No clade constraints were used in our analyses when comparing the continuous, discrete, and joint models to one another. Rather, starting trees were drawn from a distribution of trees plausible under the FBD distribution. We ran the analysis with and without sampled ancestors. For the fossil ages, we used occurrences from the PBDB, including many that were entered for the purpose of this project. Because the PBDB returns only general ages in its output, we used a separate database (see Congreve et al. 2021) that includes biozonation, radiometric dates (primarily U/Pb), and sequence stratigraphy. The estimated first-appearance age for each taxon is pulled from a uniform distribution, where the ranges are the estimated time for the most recent common ancestor for the given clade (TMRCA in `RevBayes`), minus the maximum and minimum ages. This standardizes the age of each taxon based on the clades they belong to.

One possible concern is that some of the taxa analyzed here are themselves zone taxa. For example, the *Dicynodon* zone (= *Dicynodon-Theriognathus* Subzone of the *Daptocephalus* Assemblage Zone) is thought to span from 255.2 to 254.0 Ma (Smith et al., 2020). Simply assuming that all *Dicynodon* occurrences fall within that span would present circular reasoning. However, the zone taxa analyzed here do not have a 1:1 relationship with their eponymous zones; the biozones in question represent characteristic assemblages (presence of multiple taxa simultaneously, some components of which pre- or postdate the nominal zone). Other criteria used to date PBDB sites (including U/Pb data) are sufficient to avoid possible circularity.

All the original data and scripts necessary to reproduce the analyses reported in this study can be accessed through the Dryad link: https://doi.org/10.5061/dryad.w0vt4b93r.

Clock models.— We used an uncorrelated relaxed lognormal clock to model the rate of evolution across the tree (Thorne et al. 1998; Drummond et al. 2006). This clock model is very flexible, allowing ancestor and descendant branches to have very different rates of evolution. This model was held constant across all analyses and data sets.

Joint inference.— We ran a total of five different models that include varying amounts of discrete and continuous character information (Table 1): continuous with all characters sharing a BM rate (*continuous–one-rate*), continuous with characters having their own BM rate (*continuous–multi-rate*), discrete characters under the Mk model and continuous characters in which all continuous characters share a BM rate (*combined–one-rate*),

TABLE 1. A table demonstrating the various character models tested in the paper

Name	Discrete characters included	Continuous characters included	Number of σ Brownian motion rates	Generations to converge
Continuous–multi-rate	Yes	No	23	135,000
Continuous–one-rate	Yes	No	1	235,000
Combined–one-rate	Yes	Yes	1	200,000
Combined–multi-rate	Yes	Yes	23	135,000
Discrete	Yes	No	NA	260,000

and discrete and continuous characters with continuous characters having their own BM rate (*combined–multi-rate*), all of which were compared with the discrete-only (*discrete*) model.

We performed phylogenetic inferences in `RevBayes` (Höhna 2014; Höhna et al. 2016), which allows for easily customizable scripting and extensive researcher control over parameterization. We use continuous-time Markov chain Monte Carlo (MCMC) to explore tree space (see Wright 2019 for more on the MCMC and its utility in morphological phylogenetics and Barido-Sottani et al. 2024 for a general overview of MCMC in phylogenetics). A key aspect of the MCMC is that no two analyses will be identical, but when modeling the same phenomenon one expects different analyses to converge to the same posterior distribution. Therefore, we ran two replicates of each analysis, and checked for convergence both within and between replicates. To evaluate whether or not our model has reached convergence, we use estimated sample size (ESS) as an initial marker, with an ESS of 200 being considered the threshold minimum for convergence (Lanfear et al. 2016; Rambaut et al. 2018). We checked convergence of the tree itself in `RWTY` (Lanfear et al. 2016).

We generated summary maximum clade credibility (MCC) trees for each analysis using the `mccTree` command in `RevBayes`. To do this, we discarded 25% of the sample as burn-in and combined both replicates of each analysis.

Post-Analysis Processing

Following our estimations, we quantified several summary statistics from each run. In RevGadgets, we calculated tree and branch lengths for all trees in the posterior distribution (Tribble et al. 2022). We also used the software `RWTY` (Warren et al. 2016) to calculate a treespace visualization (Hillis et al. 2005; Whidden and Matsen 2015). For the treespace visualization, we recomputed our MCMC tree logs with sampled ancestor printing suppressed (but sampled ancestors still allowed in the analysis). This is because few software packages accept sampled ancestors. We sampled the chains every 100 steps, and calculated Robinson–Foulds distance (Robinson and Foulds 1981) both within each analysis type (i.e., combined–multi-rate) and between analysis types. We then plotted these to treespace using multidimensional scaling to plot 100 points per analysis type to a 2D space (Warren et al. 2016). The resultant graphic was refined using ggplot2 (Wickham 2011) for custom colors.

RESULTS

Model Results

All model and data combinations ultimately reach convergence in our experiments, as can be seen in Supplementary Figure S1. Plots on Supplementary Figure S1 are shown downsampled, as to appear on the same axis. However, in Table 1, the number of generations required for convergence is shown. Interestingly, the discrete-only data sets take the longest to converge, followed by the combined–one-rate analysis. The combined–multi-rate analysis takes the least time. It should be noted that generations in `RevBayes` run multiple moves per iteration (Höhna et al. 2014). Thus, these are not comparable with other software packages that can use continuous traits, such as `BEAST 2` (Bouckaert et al. 2014).

As shown on Figure 2a, the different treatments arrive at different origin times for the clade. Most treatments favor an origin time of between 280 and 270 million years before the present, which is consistent with the estimated age of divergence between the major therapsid subclades (Biarmosuchia, Dinocephalia, Anomodontia, and Theriodontia) based on recent analyses examining therapsid interrelationships more broadly (Matamales-Andreu et al. 2024). The discrete and combined–multi-rate data sets and models prefer a slightly younger origin time, at around 270 Ma. These ages accord with prior work, as the oldest fossil anomodont (the larger clade containing dicynodonts and their nearest relatives) in our sample, *Biseridens qilianicus*, has an estimated age of only ~265 Ma and is morphologically very similar to basal members of other therapsid subclades (particularly dinocephalians), suggesting that there was not a protracted period of divergence prior to its appearance in the record. Even with differences in the origin times preferred by different models, there is substantial overlap between the youngest estimates and older ones.

Although we can easily see the distribution of origin times, the same is not true for major groups of organisms without instituting clade constraints. Because each tree in the posterior may have slightly different arrangements of species, it is unlikely that a single node in the analysis consistently retains the same clade across

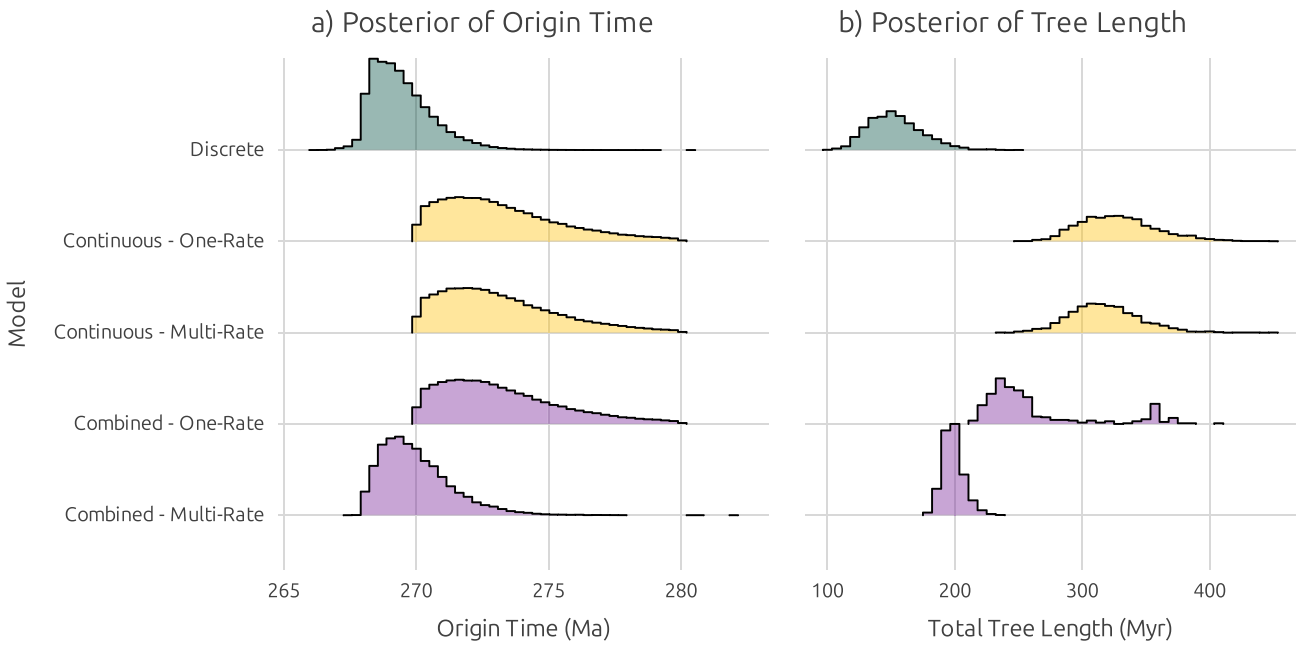


FIGURE 2. Visualization of posterior distributions of (a) origin times and (b) tree lengths (in millions of years before the present) between different models and data sources.

TABLE 2. Age estimates of maximum clade credibility trees across models compared with expected ages

Clade	Expected age (Ma)	Discrete age (Ma)	Continuous–single-rate (Ma)	Continuous–multi-rate (Ma)	Combined–single-rate (Ma)	Combined–multi-rate (Ma)
Anomodontia	274 – 270[*1]	272	275	273.5	269.5	273
Dicynodontia	269 – 265[*2]	271	272.5	271	268.5	272
Bidentalia	263[*3]	262	268.5	269.5	266.5	269
Dicynodontoidea	261[*4]	259.5	263	266	259	261
Kannemeyeriiformes	250[*5]	249	252	251	248	250

Age estimates are based on divergence dates approximated from trees depicted within the supplemental materials. Age estimates are based on figures or text in [*1]Matamales-Andreu et al. (2024), [*2], [*3-4]Angielczyk et al. (2021), [*5]Sulej and Niedźwiedzki (2019). Ages are approximations and should be viewed as having an expected error of ±1 Ma.

the entire posterior. Therefore, we can only use the MCC trees for our given treatments to evaluate how divergence dates correspond to present estimates. We checked the divergence dates across analyses for the following major clades in the taxon sample: Anomodontia, Dicynodontia, Bidentalia, Dicynodontoidea, and Kannemeyeriiformes (Table 2). The clade ages in the discrete model are largely consistent with the expected dates, albeit with Dicynodontia being younger than expected, with a rapid diversification from the Anomodontia one million years prior. Overall, the estimated ages are largely consistent with expected ages, with the most notable exceptions being the age for the Bidentalia, which was consistently recovered as older in all but the discrete analysis. This discrepancy is largely the product of members of the bidentalia being recovered among early diverging anomodonts in the continuous and combined analyses, which artificially pulls these clades closer to the age of divergence for the Dicynodontia. Outside of the age estimate for the Dicynodontoidea (specifically the continuous–multi rate model), ages are within 2–3 million years of expected divergence dates, which is a reasonable error given the inherent error expected from U/Pb radiometric dates for late Paleozoic and early Mesozoic deposits. The topologies for the continuous (both) and the combined–multi-rate analyses dissolve many of these major clades, and so ages are more reflective of earliest clade member ages, rather than the estimated age of diversification of a monophyletic lineage.

Tree Topology

Our analyses did not reproduce the consensus tree that has been produced by previous cladistic analyses and consistently supported by years of expert opinion

(Angielczyk and Kurkin 2003; Kammerer et al. 2011; Angielczyk et al. 2014; Olivier et al. 2019; Martinelli et al. 2021). We find that, in general, when continuous variables are examined alone, the produced trees are essentially nonsense, dissolving expected clades and yielding taxon groupings with minimal morphological similarity and high levels of stratigraphic debt (Fig. 3; for fully labeled trees, see Supplementary Figs. S2–S6). This is best exemplified in the Kannemeyeriiformes; this clade is known only from the Triassic Period and its members have consistently been recognized as more closely related to one another than to Permian dicynodonts (e.g., Kammerer and Ordoñez 2021). When only continuous data are incorporated, the Kannemeyeriiformes are recovered as a polyphyletic assemblage nested in two independent lineages of Permian dicynodonts. When the discrete data are examined alone, we find that Kannemeyeriiformes remains mostly intact, but a few kannemeyeriiform taxa are instead reconstructed as sister taxa to Permian dicynodonts traditionally recovered as distant to this group. Generally, however, the results of this analysis are far more similar to established morphological and stratigraphic relationships for dicynodonts than the hypotheses presented in the continuous-only data. Independently, the impact of variation in fossil record quality is evident in the characters, where expected relationships among Permian dicynodonts are more or less conserved under the discrete characters, and the continuous variables assist most strongly in determining relationships among the Triassic dicynodonts.

However, even the discrete data yield problematic taxonomic relationships at the base of Anomodontia and the dicynodont subclade Emydopoidea, specifically concerning the positions of the genera *Ulemica* and *Dicynodontoides*, respectively. *Ulemica* is a non-dicynodont anomodont, historically recovered in the clade Venyukovioidea; indeed, the primary basis for this clade's existence, as the well-preserved, extensively studied remains of this taxon were the basis for allying taxa based on more fragmentary remains such as *Venyukovia* and *Otsheria* (Ivakhnenko 1996). It should be noted that the posterior support for *Ulemica* is consistently low (< 0.04), suggesting that its placement is spurious, and should be taken lightly.

Some differences in topology relative to the "standard" parsimony-based trees are expected for our analyses; for example, although the recovered position of *Niassodon* is also problematic, this has also been a labile taxon in previous parsimony analyses of dicynodonts. Similarly, the failure to recover a monophyletic Cryptodontia is an issue that has also arisen in various prior iterations of dicynodont phylogeny (Angielczyk et al. 2017) and should not cause major concern here. By contrast, the "erroneous" position of exceptionally well-represented taxa such as *Ulemica* and *Dicynodontoides*, with extremely well-supported placements in previous trees, is highly concerning, and suggests that the explanatory power of synapomorphies in parsimony analyses does not readily transfer to Bayesian total evidence analyses.

Tree Lengths

To gain a sense of how model choice impacts phylogenetic inference, we estimate tree length as a cumulative sum of all internal branches within the tree. The discrete produces the youngest set of trees. Continuous characters (only) modeled under either the one-rate or multi-rate models favor longer trees. Combined analyses split the difference, with multi-rate models being closer in length to discrete trees. This is consistent with our findings on origin time (Fig. 2a), which reflect the continuous-only data sets favoring an older age. Overall, including continuous and discrete traits together seems to moderate the conflicting signals of age from the two trait types on their own.

As discussed in the *Fossilized birth–death model* section, we ran the analyses with and without sampled ancestors. We decided not to visualize them in our tree figures as they were numerous and poorly supported. In Supplementary Figure S7, we plot the ages of the MCMC runs with and without sampled ancestors. We find that without sampled ancestors incorporated in the model, trees grow unrealistically long. This is largely due to inflating the number of bifurcations, and therefore pushing back divergences deeper in the tree.

Treespace

When we generate a treespace (Fig. 4) comparing all five experimental treatments, we see a pattern not dissimilar from that seen in Figure 2a and 2b. The combined–multi-rate and one-rate cluster together, topologically, with the discrete characters. The continuous data types are further away in space, indicating less similar trees to the combined and discrete cluster. However, the continuous trait trees are still quite dissimilar from one another, as evidenced from their wide dispersal.

Discussion

The Unintended Consequences of Continuous Characters

Continuous variables preserve morphological characteristics via direct measurement. It would follow that such quantification would have utility in morphological phylogenetics, but we find the answer to not be so clear-cut. Inclusion of continuous characters appears to have a large effect on topological results and the manner in which treespace is explored (Fig. 4). We see this in the dissolution of clades, such as Geikiidae and Lystrosauridae, being interspersed throughout the tree when continuous variables are included (Fig. 3; for full, labeled

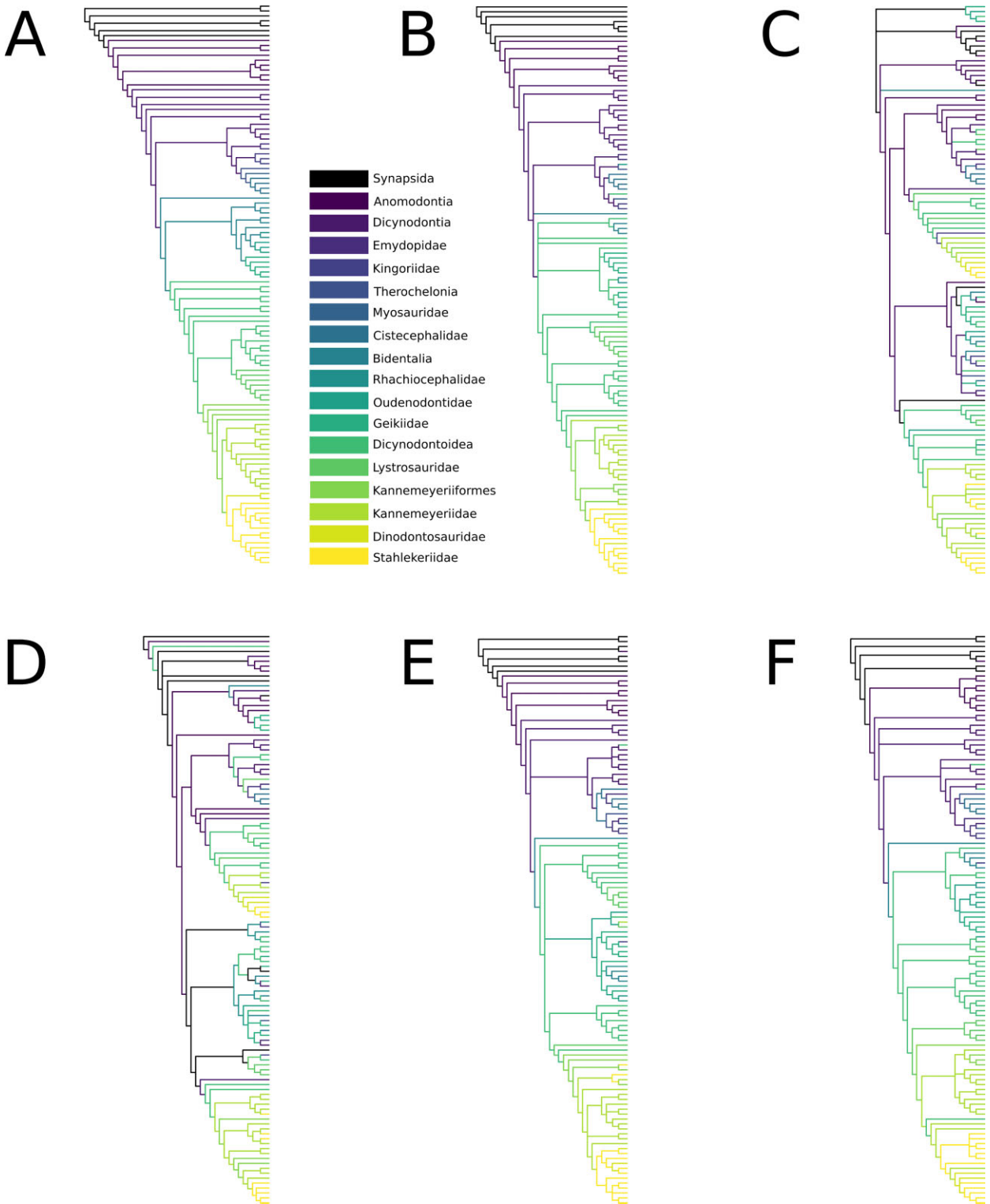


FIGURE 3. Topological differences between models plot depicting how dicynodont groupings are differently dissolved in our various analyses. (A) Current consensus topology. B–F are maximum clade credibility (MCC) trees from our five models: (B) discrete only; (C) continuous–one-rate; (D) continuous–multi-rate; (E) combined–one-rate; (F) combined–multi-rate. The currently accepted topology was drawn based on Kammerer et al. (2016, 2019); Kammerer and Ordoñez (2021). Topologies B–F were redrawn from our MCC trees. Full trees with tip annotations and clade supports are included as Supplementary Figures S2– S6. Tips for the groupings "Synapsida," "Anomodontia," "Dicynodontia," "Therochelonia," "Bidentalia," "Dicynodontoidea," and "Kannemeyeriiformes" represent the basal grades within these groups exclusive of the other listed clades.

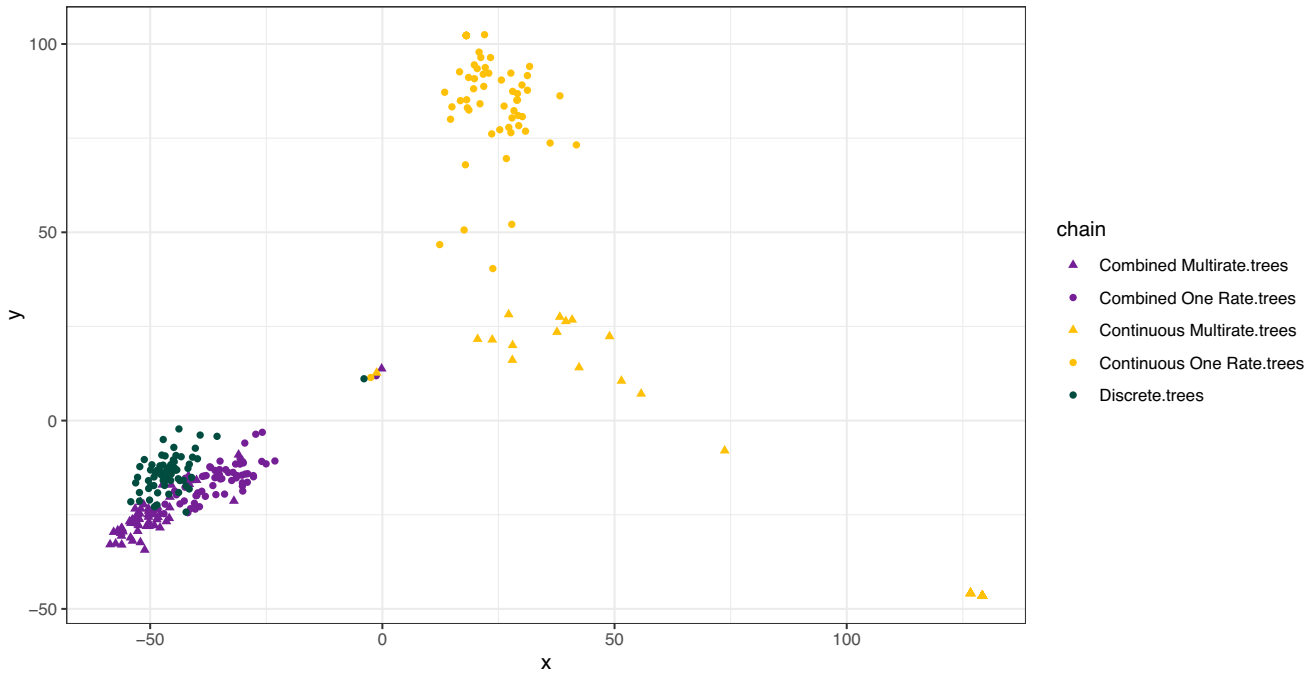


FIGURE 4. A treespace depicting how the five different experimental treatments sampled tree space. Points are colored by treatment. The metric is Robinson Foulds, which is the topological difference between two trees weighted by branch length similarity. Thus, points that are closer together are more similar topologically and in branch lengths.

trees, see Supplementary Figs. S2–S6). Geikiids are Permian dicynodonts that are short and broad-skulled, and in their slightly less inclusive subclade, Geikiinae, these taxa possess cranial ornamentation in the form of nasal and parietal bosses (Kammerer and Angielczyk 2009). The dissolution of geikiines is thus problematic because it would require numerous independent events wherein major morphological modifications to the nasal bones would need to occur; ultimately the present rarity of this characteristic in non-geikiine dicynodonts suggests homology of the nasal bosses as a better-supported hypothesis. The Lystrosauridae is clade of dicynodonts that spans the Permo-Triassic Mass Extinction event and are often regarded as globally cosmopolitan, generalist disaster taxa in earliest Triassic sediments (Viglietti et al. 2021). Although the exact position of the Lystrosauridae in greater dicynodontoid phylogeny has been poorly supported in previous analyses, the monophyly of the Lystrosauridae has been empirically supported repeatedly (Maisch 2002; Kammerer and Angielczyk 2009). These clades are well-supported in the 174 characters of the discrete-only analysis, but it only takes the addition of the 23 continuous characters to break that support. This finding is reminiscent of similar findings in combined molecular-morphological data sets (Bapst et al. 2018), in which a few morphological characters can break support for clades supported by larger volumes of molecular data. This is particularly problematic, as the continuous variables are largely used to better place the kannemeyeriiform dicynodonts and their generally fragmentary preservation when compared with Permian dicynodonts. Although there are species of kannemeyeriiform dicynodonts known from complete skeletons (e.g., *Shaanbeikannemeyeria xilougouensis*, Liu 2022; *Ishigualastia jenseni*, Cox 1965; *Kannemeyeria simocephalus*, Govender et al. 2008), much of the kannemeyeriiform record is comprised of incomplete skeletons and often fragmentary elements (e.g., *Argodicynodon boreni*, Mueller et al. 2023; *Eubrachiosaurus browni*, Kammerer et al. 2013; *Sungeodon kinkraemerae*, Maisch and Matzke 2014). Given this, diagnostic cranial features are missing in some taxa, whereas measurements and ratios can still be coded for their fragmentary postcrania. Thus, with discrete characters underrepresented in the kannemeyeriiform data, these continuous characters are expected to have a proportionally larger impact on kannemeyeriiform relationships than they would on non-kannemeyeriiform dicynodont relationships. This group is non-monophyletic when exploring the discrete-only model, and are broken up more significantly throughout the tree when continuous traits are included. The fact that these characters essentially produce a result opposite to their intent when treating them as character evolution models should raise caution concerning the use of such data in any phylogenetic context.

The complication surrounding continuous variables in phylogenetics, especially in the context of branch lengths, is potentially due to model misspecification. When examining our trees, we find what appear to be opposing results when continuous characters are treated

under a single rate or individual rates for each continuous character, particularly when in combination with discrete characters (Fig. 3). This suggests that a single-rate BM model is likely to be inappropriate for all characters. This is perhaps unsurprising, as we do not assume all discrete characters evolve at the same rate (Yang 1994; Wright and Wynd 2024). We also see this impact in the parameters of the FBD model—when we use a multi-rate model for combined discrete and continuous data sets, the origin time aligns very well with the known fossil record for the group (Fig. 2a; Fröbisch 2009). This largely also accords with the discrete-only age.

However, when only a single rate is imposed, the model struggles to find reasonable topologies that satisfy each and every continuous variable sharing the same history. When we incorporate discrete characters, it is simpler to include the continuous characters with a one-rate BM. The model has less difficulty finding starting parameters that are computable. This is likely due to the simplicity of the model space as a result of only adding one parameter. When the multi-rate BM model is implemented, 23 separate parameters need to be adjusted each iteration, increasing model complexity and computation time. However, despite this complexity, the multi-rate model reaches convergence in far fewer generations once it begins running (Table 1). As we will discuss in the *Insights Based on Dicynodont Evolution* section, the multi-rate models produce more reasonable estimates along several fronts, and we recommend exploring them in analyses incorporating continuous traits, a finding that has recently also been reported in Wright and Hopkins (2025), in which the authors corroborate out finding that continuous and discrete analyses often search treespace in very different ways.

Character evolution models are extremely powerful tools when the model of evolution properly matches the history of the trait. Existing character evolution models include those that explore random walks of evolution (i.e., BM), where the inherent assumption is that this is a trait that evolved largely via genetic drift. This can be a serious issue if the evolutionary history of the trait is non-Brownian, or if the evolution is directional (i.e., OU character evolution models; Hansen 1997; Butler and King 2004). This kind of character model is more appropriate when applied to a trait that appears to have an optimum (θ). This is not to say that there is a true optimum, but that there was a selective pressure on that trait, which eventually reached a point of general stability. The OU model assumes rapid evolution early in the trait history followed by relatively slow change once the optimum is reached. BM is a memory-rich process where the current state is dependent on the state prior to it. With stochastic change centered around a mean of 0, the model depends on small differences or long branches to accumulate changes, such that the root state of the tree is informative for the tip state of the traits. OU models, however, are memory rich in the stability phase, but large shifts during the directional evolution phase suggest that the root state is less useful when estimating characters that do not evolve by random evolutionary processes. OU processes may be a better way to model continuous traits for phylogenetic inference, but are not currently implemented for inference to test.

Our analyses with continuous traits still produce interesting artifacts. As seen on Figure 2b, continuous traits run under either multi-rate or one-rate BM processes produce much longer tree lengths than either discrete or combined discrete-continuous data sets. Conceptually, this makes sense as to why it would be a problem in the Bayesian analysis, but not the cladistic analysis *sensu* Farris (1970). Because the additive method only requires tips, and nodes, the differences between the tips and nodes would be specific to each individual character, such that characters with vastly distinct values and rates of change can be independently estimated using the same algorithm. Thus, the additive method of Farris has more in common with the multi-rate than one-rate BM. In the case of strict BM, the focus on branch lengths means that a single rate parameter is unlikely to properly explain the evolutionary history of completely different traits, let alone, 23 different traits encompassing different skeletal elements.

Potentially the most impactful biases of the fossil record is preservation (Webster and Hughes 1999). Although we have thus far extolled the objectivity of continuous trait data, they are not without biases. For both the vertebrate and invertebrate fossil records, specimens are rarely preserved as they would have been in life. Whether it be soft tissues that are inherently unlikely to preserve, broken or partially preserved elements, entirely missing elements, or distortions from shifts to the local geography, fossilization obliterates and obscures the biological entities it captures. Although continuous trait measures may have less of a human interpretive filter, raw measurements taken directly from fossil specimens are rarely to be trusted without some degree of scrutiny. Common approaches to accounting for taphonomy—the process of fossilization—include retrodeformation analyses (Angielczyk and Sheets 2007; Arbour and Currie 2012; Lautenschlager 2016; Schlager et al. 2018; Kammerer et al. 2020), extrapolations based on numerous specimens (Lande 1979; Alexander 1985; Klingenberg 1996; Klingenberg et al. 2001), and our own work using linear mixed-effects models to account for random deformation (Wynd et al. 2021). These methods are potentially of use to researchers looking to use continuous traits inferentially.

However, each of these approaches carries their own pitfalls that limit broad application for small data sets, which account for the vast majority of existing paleontological data sets (Koch 1991). Because raw measurements of fossils still include significant bias, phylogenetic characters have been primarily based on observations that can be discretized (Baum 1988). It is clear that discretized characters will always be a staple of phylo-

genetic analyses based on fossils; regardless, the incorporation of continuous variables is critical to maximize the amount of data that can be brought to bear on a problem.

The Benefits of Including Continuous Variables in Joint Phylogenetic Estimation

The issues pointed out above bear on the signal within continuous characters and how we need to thoroughly vet both our data and how it behaves under differing model circumstances. As seen in Figure 2a and b, continuous traits can have a powerful impact on the parameters of our models, and on the way treespace explored (Fig. 4). If continuous traits had no signal, we would not observe these differences. Therefore, it is crucial to think carefully about the choices used to model them. When only continuous characters are considered, trees are much longer (Fig. 2b) than when data sets are combined or discrete. Importantly, when more than a single rate is specified for the Brownian model in a combined data set, there is no major change to the origin time, and branch lengths are consistent with those generated through discrete characters under an Mk model (Fig. 2a and b).

We discuss above how, aside from branch lengths, the additive model is conceptually similar to a character evolution model following multi-rate BM. However, the additive model as a parsimony-based method lacks a unifying rate parameter, thus the additive model can effectively incorporate characters with lots of rate heterogeneity. The BM character evolution model, on the hand, requires some value to ascribe to the rate of change for the trait in question; and when only a single rate is given for over 20 characters, many of which sample completely different skeletal elements, it is unlikely that those characters are going to contribute meaningful information, and will instead produce erroneous clades. This may seem discouraging, but the opposite is actually true. It should be very validating that the model fails so catastrophically when we try to shoehorn fundamentally distinct data into the same parameters and expect the model to appropriately sample the nuance. This might be considered similarly to not allowing among-character rate variation in a substitution model (Yang 1994; Wright and Wynd 2024; Capobianco and Höhna 2025). The addition of discrete traits as a separate partition substantially helps inference in both the one- and multi-rate models.

Insights Based on Dicynodont Evolution

When it comes to dicynodont topology, the discrete and combined-strict analyses produce topologies most consistent with the currently supported dicynodont relationships. For the majority of Permian dicynodonts, the historically recognized clades are largely preserved and the species based on more complete specimens tend to be placed in their expected positions. However, the placement of some taxa should raise concerns. Specifically, the genera *Ulemica* and *Dicynodontoides* are some of the better sampled taxa in our data set, and in parsimony analyses, their positions are supported by minimally dozens of synapomorphies, yet their phylogenetic positions in our analyses are highly variable and are supported by extremely low posterior probabilities. *Ulemica* is recovered well-nested within the Dicynodontia in the discrete-only analysis, but is pulled closer to its "proper" position when continuous characters are incorporated; however, the continuous-only analyses create entirely unrealistic topologies in the rest of the tree. The recovery of an "erroneous" (relative to established phylogenies) position for *Ulemica* is notable because it is known from a reasonably complete skull allowing it to be coded for the majority of characters in the analysis. Its position in parsimony-based analyses as a taxon falling outside of Dicynodontia is supported by 33 discrete dicynodont synapomorphies not present in *Ulemica* (based on analysis of Kammerer et al. 2019). Given the strong support in previous analyses, it is both unlikely and strange that *Ulemica* would behave as a rogue taxon with such low stability between analyses, especially compared with taxa known from much more fragmentary material. The case for *Dicynodontoides* should be even more concerning, as this taxon is one of the most complete taxa sampled in our analyses, being scored for over 95% of characters, including extensively studied postcranial skeletons (Angielczyk et al. 2009). Typically a non-bidentalian emydopoid, it is recovered here as belonging to Bidentalia (discrete-only and combined–multi-rate), and in some analyses more specifically among lystrosaurids (continuous only one-rate and multi-rate) or as sister taxon to geikiids (combined–one-rate). These latter two groups are supported by numerous synapomorphies in previous analyses that are not present in *Dicynodontoides*, and in no prior analysis has it been recovered as closely related to them (pers. obs. C.F.K.).

Given the completeness of these species, we lack any strong arguments as to why these taxa seem to have so much lability; it is possible that the explanatory power of synapomorphies in parsimony analysis does not appropriately transfer to Bayesian total-evidence analyses, but this will require more targeted studies to address at a broader scale. Outside of *Ulemica* and *Dicynodontoides*, the Bayesian discrete and combined analyses tend to largely support the expected clades, with most other rogue taxa having a known record of phylogenetic lability under parsimony analyses. Because the driving force behind the use of continuous characters in this data set is to better place the Kannemeyeriiformes, these results make intuitive sense. When discrete characters are included, in general, the species known from the most complete specimens are in or near their expected phylogenetic positions. With the continuous characters however, the impacts of model misspecification are directly represented on the phylogeny via the "incorrect"

placements of kannemeyeriiform taxa. Although both the strict and multi-rate continuous-only models produce trees that are effectively nonsense, the multi-rate model tends to preserve smaller clades and intersperse them throughout the tree in seemingly random clades. The strict model on the other hand tends to take more individual species and intersperse those throughout the tree. We interpret this difference as the strict model attempting to "shoehorn" characters that likely evolved at a different rate from one another into a single explanatory rate; this is an issue that is not present in the parsimony analysis because the additive character model of popular tree-building software TNT is based on minimizing the between-node differences across the tree rather than attempting to apply a rate to a branch length.

In the divergence dates, the different analyses produce dates more or less in line with prior work. Each of the analyses produces an age for Anomodontia that are largely consistent with recent tip-dated and time-calibrated phylogenies (Sulej and Niedźwiedzki 2019; Angielczyk et al. 2021; Matamales-Andreu et al. 2024). Except for the combined–multi-rate analysis, the age of the Dicynodontia is consistently older than would be expected, and is largely driven by non-anomodontian therapsids being interspersed throughout early anomodonts and dicynodonts, artificially pulling the age of Dicynodontia back. The same can be seen in the Bidentalia, but only the discrete analysis is consistent with the expected age. This is largely driven by the dissolution of the Cryptodontia, the sister taxon to the Dicynodontoidea. Across analyses, major groups of Cryptodonts are largely retained (e.g., the Oudenodontidae, Rhachiocephalidae, and Geikiidae), but these clades are only joined in a common clade in the discrete analysis. Geikiids tend to be pulled into the Dicynodontoidea as their earliest diverging members, which would cause the divergence date for the clade to be overestimated. A similar pattern is seen in the Kannemeyeriiformes, where kannemeyeriids and stahleckeriids are not found as sister taxa, but are frequently nested within subsets of dicynodontoids, a hypothesis that has not previously been recovered cladistically and is also not supported by the existing taxonomic literature of dicynodonts. Ultimately, the repeated overestimates for clade divergence dates are driven by older taxa and sometimes entire clades being recovered as nested in younger clades, usually what is expected to be their sister group.

For origin times, the discrete and combined–multi-rate models produce the most reasonable results given prior work. Although the other analyses produce results that are somewhat older (Fig. 2a), they still have considerable overlap. The more striking concerns in the taxon ages are exemplified in the temporal disconnect between the two species of *Eodicynodon*. Although both "*Eodicynodon*" (now *Nyaphulia*) *oelofseni* and *Eodicynodon oosthuizeni* were found at the same locality, mere meters away from one another, only the combined–multi-rate analysis recovers these species as synchronous. This is most likely due to the ways in which we collect paleontological data, and the lack of autapomorphic characters in standard phylogenetic inference in paleontology. Because true autapomorphies, traits that are only present in terminal tips, do not influence the results of a parsimony analysis, they are generally described within the text, but very seldom are they coded into and included in the phylogenetic character matrix. This lack of coded autapomorphies means that the characters that differentiate tips are dominated by reversals or local autapomorphies. In the case of *Eodicynodon*, in our temporally distinct analyses we recover *E. oosthuizeni* as the sister taxon to *Lanthanostegus mohoii*. This relationship is important because it would require multiple reversals of synapomorphies in *E. oosthuizeni* to produce such a relationship. The additional local autapomorphies in *E. oosthuizeni* offer an explanation for the differing ages reconstructed in otherwise synchronous species. It is likely that the inclusion of additional autapomorphies would help provide better age constraints in Bayesian analyses (Matzke and Irmis 2018). The most important part of our models are their failures, which make perfect sense given the nature of paleontological data and the present limits around using character evolution models in phylogenetic estimation. Perhaps the most important note for paleontologists is to include autapomorphies in character matrices; to not only incorporate autapomorphies into the same matrices as, and context in which we identify synapomorphies, but also to allow for future analyses and comparisons between parsimony and tip-dated phylogenetic inference. This is, additionally, an interesting consideration for methods developers to consider the role of lagerstattë effects in the FBD model.

An Objective Future for Fossil Phylogenetics

Molecular phylogenetics has long established incorporating multiple forms of information for joint estimation, sometimes labeled as "total evidence" analysis. Over the years this has often incorporated both molecular, morphological, and sample age evidence (Heath et al. 2014). But it has more seldomly included continuous trait evidence (Parins-Fukuchi 2018; Álvarez-Carretero et al. 2019; Zhang et al. 2024). In the field of comparative methods, there are established numerous continuous character evolution models, including BM and its variants, and the ways that they evolve along a phylogenetic tree. Joint inference using character evolution models is a critical step forward in Bayesian phylogenetics, as any number of data types can be included flexibly by expanding the model graph (Warnock and Wright 2020). This allows continuous variables to be included and used to explore distinct regions of trait space that may be otherwise uninvestigated. Uniting the simplest character evolution model (BM) along with the simplest morphological model (Mk, Lewis 2001) estab-

lishes a critical starting point for advancing phylogenetic inference in the paleobiological sciences. Further questions, such as how character correlations and/or dependency affect topology, can be usefully asked in this framework.

Likelihood-based phylogenetics has historically attempted to incorporate continuous variables through a variety of models. These include the threshold models, which discretize continuous variables that have numerical values exceeding a threshold (Felsenstein 1988), and OU models, which have trait optima (Hansen 1997). These models have mostly been used for *post hoc* trait evolution inference. This represents some of the earliest work in incorporating character evolution models into the estimation of a phylogenetic tree (but see Parins-Fukuchi 2018; Zhang et al. 2024) using a data set for which both continuous and discrete traits are available, enabling direct signal comparison. In both of these studies, continuous traits were found to be a significant boon for phylogenetics, due to their relatively lower evolutionary rates and lack of homoplasy. Although this study has highlighted some complications to the incorporation of continuous traits, we have also fundamentally found that these traits are contributing meaningful signal that we can model to improve our phylogenetic estimations.

The incorporation of continuous variables will be a net benefit to all life sciences, but will have a disproportionately large impact in paleontological communities, where information is limited due simply to the process of fossilization. These joint inference models provide flexibility in the kinds of data and the way those data are treated, which in turn provides the investigator more information in an otherwise limited framework. As model choice appears to have a very strong impact on tree inference and the ability to reach convergence, it is inevitable that additional character evolution models, such as the OU model, which is meant to be an approximation of selection, will be necessary to obtain our best approximations of phylogeny. Additionally, among-character correlations may be an important next step in this work. Developmental processes and modular evolution (Clarke and Middleton 2008) can cause suites of traits to be non-independent of one another, a fact that likely affects both discrete and continuous traits. This problem has received relatively little attention in Bayesian analyses, but may lead to an overrepresentation of some signals within the data.

The field of paleontology has a long history of co-opting neontological tools and ideas, usually a decade or two after original publication, for use with fossils. This has led to paradigm shifts within the field, including ecological abundance indices, CT scanning, and in our opinion, Bayesian phylogenetics is likely to be a future paradigm shift in the field. Joint inference integrates neontological concepts and tools, such as clock models, with modern quantitative paleobiological tools, such as the fossilized birth–death process. The useful synthesis of paleontological and neontological ways of viewing the world has provided a unique framework for testing the impact of different character evolution models on phylogenetic estimation. Core issues remain, such as that the direct quantification of morphology tends to either obscure nuance when changes are textural, or they require high-resolution 3D scans accompanied by advanced computational methods that can produce and compare topological maps of morphology (Boyer 2008; Goswami et al. 2014; Fabre et al. 2020). As we propose in the *The Unintended Consequences of Continuous Characters* section, there are numerous ways forward to deal with these issues. We hope this work will form a basis for interesting and productive work on the statistical problems associated with continuous trait incorporation in phylogenetic analyses.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.[NNNN]

Funding

A.M.W., B.M.W., and B.K. were supported on NSF DEB 2045842 and A.M.W. was supported on NSF CIBR 2113425. A.M.W. additionally received funds from Southeastern Louisiana University's Dyson Fellowship. P.J.W. was supported on NSF EAR 2129628. B.K. was additionally supported by the European Union (ERC, MacDrive, GA 101043187).

Acknowledgments

Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. Finally, we thank associate editor Dr. Joëlle Barido-Sottani, reviewer Dr. Caroline Abbott, and two anonymous reviewers for their incisive and helpful comments.

Data Availability

All the original data and scripts necessary to reproduce the analyses reported in this study can be accessed through the Dryad link: https://doi.org/10.5061/dryad.w0vt4b93r

References

Alexander R.M. 1985. Chapter 2. Body support, scaling, and allometry. Functional vertebrate morphology. Harvard University Press. p. 26–37.

Álvarez-Carretero S., Goswami A., Yang Z., Dos Reis M. 2019. Bayesian estimation of species divergence times using correlated quantitative characters. Syst. Biol. 68(6):967–986. https://doi.org/10.1093/sysbio/syz015

Angielczyk K.D., Hancox P.J., Nabavizadeh A. 2017. A redescription of the Triassic kannemeyeriiform dicynodont sangusaurus (Therapsida, Anomodontia), with an analysis of its feeding system. J. Vertebr. Paleontol. 37(Suppl 1):189–227. https://doi.org/10.1080/02724634.2017.1395885

Angielczyk K.D., Kurkin A.A. 2003. Phylogenetic analysis of Russian Permian dicynodonts (Therapsida: Anomodontia): implications for Permian biostratigraphy and Pangaean biogeography. Zool. J. Linnean Soc. 139(2):157–212. https://doi.org/10.1046/j.1096-3642.2003.00081.x

Angielczyk K.D., Liu J., Yang W. 2021. A redescription of *Kunpania scopulusa*, a bidentalian dicynodont (Therapsida, Anomodontia) from the ?Guadalupian of northwestern China. J. Vertebr. Paleontol. 41(1):e1922428. https://doi.org/10.1080/02724634.2021.1922428

Angielczyk K.D., Sheets H.D. 2007. Investigation of simulated tectonic deformation in fossils using geometric morphometrics. Paleobiology 33(1):125–148. https://doi.org/10.1666/06007.1

Angielczyk K.D., Sidor C.A., Nesbitt S.J., Smith R.M., Tsuji L.A. 2009. Taxonomic revision and new observations on the postcranial skeleton, biogeography, and biostratigraphy of the dicynodont genus *Dicynodontoides*, the senior subjective synonym of *Kingoria* (Therapsida, Anomodontia). J. Vertebr. Paleontol. 29(4):1174–1187. https://doi.org/10.1671/039.029.0427

Angielczyk K.D., Steyer J.S., Sidor C.A., Smith R.M., Whatley R.L., Tolan S. 2014. Permian and Triassic dicynodont (Therapsida: Anomodontia) faunas of the Luangwa Basin, Zambia: taxonomic update and implications for dicynodont biogeography and biostratigraphy. In: Kammerer, C., Angielczyk, K., Fröbisch, J., editors. Early evolutionary history of the Synapsida. Vertebrate Paleobiology and Paleoanthropology. Dordrecht: Springer. p. 93–138.

Arbour V.M., Currie P.J. 2012. Analyzing taphonomic deformation of ankylosaur skulls using retrodeformation and finite element analysis. PloS One 7(6):e39323. https://doi.org/10.1371/journal.pone.0039323.

Bapst D.W., Schreiber H.A., Carlson S.J. 2018. Combined analysis of extant Rhynchonellida (Brachiopoda) using morphological and molecular data. Syst. Biol. 67(1):32–48. https://doi.org/10.1093/sysbio/syx049

Bardin J., Rouget I., Cecca F. 2017. Ontogenetic data analyzed as such in phylogenies. Syst. Biol. 66(1):23–37.

Barido-Sottani J., Justison J.A., Wright A.M., Warnock R.C., Pett W., Heath T.A. 2020. Estimating a time-calibrated phylogeny of fossil and extant taxa using RevBayes. In: Scornavacca C., Delsuc F., Galtier N., editors. Phylogenetics in the genomic era. No commercial publisher—authors open access book. p. 5.2:1–5.2:23.

Barido-Sottani J., Schwery O., Warnock R.C., Zhang C., Wright A.M. 2024. Practical guidelines for Bayesian phylogenetic inference using Markov Chain Monte Carlo (MCMC). Open Res. Europe 3(204):204. https://doi.org/10.12688/openreseurope.16679.3

Baum B.R. 1988. A simple procedure for establishing discrete characters from measurement data, applicable to cladistics. Taxon 37(1):63–70. https://doi.org/10.2307/1220934

Beaulieu J.M., Jhwueng D.C., Boettiger C., O'Meara B.C. 2012. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. Evolution 66(8):2369–2383. https://doi.org/10.1111/j.1558-5646.2012.01619.x

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10(4):e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Boyer D.M. 2008. Relief index of second mandibular molars is a correlate of diet among prosimian primates and other euarchontan mammals. J. Hum. Evol. 55(6):1118–1137. https://doi.org/10.1016/j.jhevol.2008.08.002

Brazeau M.D. 2011. Problematic character coding methods in morphology and their effects. Biol. J. Linn. Soc. 104(3):489–498. https://doi.org/10.1111/j.1095-8312.2011.01755.x

Butler M.A., King A.A. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164(6):683–695. https://doi.org/10.1086/426002

Capobianco A., Höhna S. 2025. On the Mkv model with among-character rate variation. Syst. Biol. syaf038. https://doi.org/10.1093/sysbio/syaf038

Chappill J.A. 1989. Quantitative characters in phylogenetic analysis. Cladistics 5(3):217–234. https://doi.org/10.1111/j.1096-0031.1989.tb00487.x

Clarke J.A., Middleton K.M. 2008. Mosaicism, modules, and the evolution of birds: results from a Bayesian approach to the study of morphological evolution using discrete character data. Syst. Biol. 57(2):185–201. https://doi.org/10.1080/10635150802022231

Congreve C.R., Patzkowsky M.E., Wagner P.J. 2021. An early burst in brachiopod evolution corresponding with significant climatic shifts during the great ordovician biodiversification event. Proc. R. Soc. B. Biol. Sci. 288:20211450. https://doi.org/10.1098/rspb.2021.1450

Cox C.B. 1965. New Triassic dicynodonts from South America, their origins and relationships. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 248(753):457–514.

Damiani R., Vasconcelos C., Renaut A., Hancox J., Yates A. 2007. *Dolichuranus primaevus* (Therapsida: Anomodontia) from the middle Triassic of Namibia and its phylogenetic relationships. Palaeontology 50(6):1531–1546. https://doi.org/10.1111/j.1475-4983.2007.00727.x

Drummond A., Ho S., Phillips M., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4(5):e88. https://doi.org/10.1371/journal.pbio.0040088

Fabre A.C., Bardua C., Bon M., Clavel J., Felice R.N., Streicher J.W., Bonnel J., Stanley E.L., Blackburn D.C., Goswami A. 2020. Metamorphosis shapes cranial diversity and rate of evolution in salamanders. Nat. Ecol. Evol. 4(8):1129–1140. https://doi.org/10.1038/s41559-020-1225-3

Farris J.S. 1970. Methods for computing Wagner trees. Syst. Biol. 19(1):83–92. https://doi.org/10.1093/sysbio/19.1.83

Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25(5):471–492.

Felsenstein J. 1985a. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39(4):783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Felsenstein J. 1985b. Phylogenies and the comparative method. Am. Nat. 1–15. https://doi.org/10.1086/284325.

Felsenstein J. 1988. Phylogenies and quantitative characters. Ann. Rev. Ecol. Syst. 445–471.

Fiorillo A.R., Padian K., Musikasinthorn C. 2000. Taphonomy and depositional setting of the Placerias quarry (Chinle formation: late Triassic, Arizona). Palaios 15(5):373–386. https://doi.org/10.1669/0883-1351(2000)015⟨0373:TADSOT⟩2.0.CO;2

Fröbisch J. 2009. Composition and similarity of global anomodont-bearing tetrapod faunas. Earth Sci. Rev. 95(3–4):119–157. https://doi.org/10.1016/j.earscirev.2009.04.001

Gingerich P.D. 1993. Quantification and comparison of evolutionary rates. Am. J. Sci. 293(A):453. https://doi.org/10.2475/ajs.293.A.453

Goloboff P.A., Mattoni C.I., Quinteros A.S. 2006. Continuous characters analyzed as such. Cladistics 22(6):589–601. https://doi.org/10.1111/j.1096-0031.2006.00122.x

Goswami A., Smaers J.B., Soligo C., Polly P.D. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. Philos. T. R. Soc. B Biol. Sci. 369(1649):20130254. https://doi.org/10.1098/rstb.2013.0254

Govender R., Hancox P.J., Yates A.M. 2008. Re-evaluation of the postcranial skeleton of the Triassic dicynodont Kannemeyeria simocephalus from the Cynognathus Assemblage Zone (Subzone b) of South Africa. Palaeontologia Africana 43:19–37.

Griffin C.T., Stocker M.R., Colleary C., Stefanic C.M., Lessner E.J., Riegler M., Formoso K., Koeller K., Nesbitt S.J. 2021. Assessing ontogenetic maturity in extinct saurian reptiles. Biol. Rev. 96(2):470–525. https://doi.org/10.1111/brv.12666

Hansen T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51(5):1341–1351. https://doi.org/10.1111/j.1558-5646.1997.tb01457.x

Heath T.A., Huelsenbeck J.P., Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. Proc.

Natl. Acad. Sci. USA 111(29):E2957–E2966. https://doi.org/10.1073/pnas.1319091111
Hillis D.M., Heath T.A., John K.S. 2005. Analysis and visualization of tree space. Syst. Biol. 54(3):471–482. https://doi.org/10.1080/10635150590946961
Höhna S. 2014. Likelihood inference of non-constant diversification rates with incomplete taxon sampling. PloS One 9(1):e84184. https://doi.org/10.1371/journal.pone.0084184
Höhna S., Heath T.A., Boussau B., Landis M.J., Ronquist F., Huelsenbeck J.P. 2014. Probabilistic graphical model representation in phylogenetics. Syst. Biol. 63(5):753–771. https://doi.org/10.1093/sysbio/syu039
Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol. 65(4):726–736. https://doi.org/10.1093/sysbio/syw021
Hoyal Cuthill J.F. 2015. The morphological state space revisited: what do phylogenetic patterns in homoplasy tell us about the number of possible character states? Interface Focus 5(6):20150049. https://doi.org/10.1098/rsfs.2015.0049
Huelsenbeck J.P., Rannala B. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. Evolution 57(6):1237–1247. https://doi.org/10.1111/j.0014-3820.2003.tb00332.x
Hunt G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. Proc. Natl. Acad. Sci. USA 104(47):18404–18408. https://doi.org/10.1073/pnas.0704088104
Ivakhnenko M.F. 1996. Primitive anomodonts, venyukoviids, from the Late Permian of Eastern Europe. Paleontolog. J. 30(5):575–582.
Jardine N. 1969. The observational and theoretical components of homology: a study based on the morphology of the dermal skull-roofs of rhipidistian fishes. Biol. J. Linn. Soc. 1(4):327–361. https://doi.org/10.1111/j.1095-8312.1969.tb00125.x
Jukes T., Cantor C. 1969. Evolution of protein molecules. In: Munro H.N., editor. Mammalian protein metabolism. New York: Academic Press.p. 21–132. https://doi.org/10.1016/B978-1-4832-3211-9.50009-7
Kammerer C.F., Angielczyk K.D. 2009. A proposed higher taxonomy of anomodont therapsids. Zootaxa 2018(1):1–24. https://doi.org/10.11646/zootaxa.2018.1.1
Kammerer C.F., Angielczyk K.D., Fröbisch J. 2011. A comprehensive taxonomic revision of *Dicynodon* (Therapsida, Anomodontia) and its implications for dicynodont phylogeny, biogeography, and biostratigraphy. J. Vertebr. Paleontol. 31(Suppl 1):1–158. https://doi.org/10.1080/02724634.2011.627074
Kammerer C.F., Bandyopadhyay S., Ray S. 2016. A new taxon of cistecephalid dicynodont from the upper Permian Kundaram Formation of India. Pap. Palaeontol. 2(4):569–584. https://doi.org/10.1002/spp2.1055
Kammerer C.F., Deutsch M., Lungmus J.K., Angielczyk K.D. 2020. Effects of taphonomic deformation on geometric morphometric analysis of fossils: a study using the dicynodont diictodon feliceps (Therapsida, Anomodontia). PeerJ 8: e9925. https://doi.org/10.7717/peerj.9925
Kammerer C.F., Fröbisch J., Angielczyk K.D. 2013. On the validity and phylogenetic position of *Eubrachiosaurus browni*, a kannemeyeriiform dicynodont (Anomodontia) from Triassic North America. PloS One 8(5):e64203. https://doi.org/10.1371/journal.pone.0064203
Kammerer C.F., Ordoñez M.d.l.A. 2021. Dicynodonts (Therapsida: Anomodontia) of South America. J. South Amer. Earth Sci. 108:103171. https://doi.org/10.1016/j.jsames.2021.103171
Kammerer C.F., Viglietti P.A., Hancox P.J., Butler R.J., Choiniere J.N. 2019. A new kannemeyeriiform dicynodont (*Ufudocyclops mukanelai*, gen. et sp. nov.) from Subzone C of the *Cynognathus* Assemblage Zone, Triassic of South Africa, with implications for biostratigraphic correlation with other African Triassic faunas. J. Vertebr. Paleontol. 39(2):e1596921. https://doi.org/10.1080/02724634.2019.1596921
Khakurel B., Grigsby C., Tran T.D., Zariwala J., Höhna S., Wright A.M. 2024. The fundamental role of character coding in Bayesian morphological phylogenetics. Syst. Biol. 73(5):861–871. https://doi.org/10.1093/sysbio/syae033.
Klingenberg C.P. 1996. Multivariate allometry. Advances in morphometrics. Springer. p. 23–49.
Klingenberg C.P., Badyaev A.V., Sowry S.M., Beckwith N.J. 2001. Inferring developmental modularity from morphological integration: analysis of individual variation and asymmetry in bumblebee wings. Am. Nat. 157(1):11–23. https://doi.org/10.1086/317002
Klopfstein S., Ryer R., Coiro M., Spasojevic T. 2019. Mismatch of the morphology model is mostly unproblematic in total-evidence dating: insights from an extensive simulation study. bioRxiv. https://doi.org/10.1101/679084.
Klopfstein S., Vilhelmsen L., Ronquist F. 2015. A nonstationary Markov model detects directional evolution in hymenopteran morphology. Syst. Biol. 64(6):1089–1103. https://doi.org/10.1093/sysbio/syv052
Koch C.F. 1991. Sampling from the fossil record. Short Courses Paleontol. 4:4–18. https://doi.org/10.1017/S2475263000002099
Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution 402–416.
Lanfear R., Hua X., Warren D.L. 2016. Estimating the effective sample size of tree topologies from Bayesian phylogenetic analyses. Genome Biol. Evol. 8(8):2319–2332. https://doi.org/10.1093/gbe/evw171
Lautenschlager S. 2016. Reconstructing the past: methods and techniques for the digital restoration of fossils. R. Soc. Open Sci. 3(10):160342. : https://doi.org/10.1098/rsos.160342
Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50(6):913–925. https://doi.org/10.1080/106351501753462876
Liu J. 2022. On kannemeyeriiform dicynodonts from the Shaanbeikannemeyeria Assemblage Zone of the Ordos Basin, China. Vert. Palasiatica 60(3):212–248.
Maisch M.W. 2002. A new basal lystrosaurid dicynodont from the Upper Permian of South Africa. Palaeontology 45(2):343–359. https://doi.org/10.1111/1475-4983.00240
Maisch M.W., Matzke A.T. 2014. Sungeodon kimkraemerae n. gen. n. sp., the oldest kannemeyeriiform (Therapsida, Dicynodontia) and its implications for the early diversification of large herbivores after the P/T boundary. Neues Jahrb. Geol. Paläontol. Abh. 272:1–12. https://doi.org/10.1127/0077-7749/2014/0394
Martinelli A.G., Escobar J.A., Francischini H., Kerber L., Müller R.T., Rubert R., Schultz C.L., Da-Rosa A.A. 2021. New record of a stahleckeriid dicynodont (Therapsida, Dicynodontia) from the Late Triassic of southern Brazil and biostratigraphic remarks on the Riograndia Assemblage Zone. Hist. Biol. 33(11):3101–3110. https://doi.org/10.1080/08912963.2020.1850715
Matamales-Andreu R., Kammerer C.F., Angielczyk K.D., Simões T.R., Mujal E., Galobart A., Fortuny J. 2024. Early-middle Permian Mediterranean gorgonopsian suggests an equatorial origin of therapsids. Nat. Commun. 15(10346):1–12. https://doi.org/10.1038/s41467-024-54425-5
Matzig D.N., Marwick B., Riede F., Warnock R.C. 2024. A macroevolutionary analysis of European late upper palaeolithic stone tool shape using a Bayesian phylodynamic framework. R. Soc. Open Sci. 11(8):240321. https://doi.org/10.1098/rsos.240321
Matzke N.J., Irmis R.B. 2018. Including autapomorphies is important for paleontological tip-dating with clocklike data, but not with non-clock data. PeerJ 6:e4553. https://doi.org/10.7717/peerj.4553
May M.R., Rothfels C.J. 2023. Diversification models conflate likelihood and prior, and cannot be compared using conventional model-comparison tools. Syst. Biol. 72(3):713–722. https://doi.org/10.1093/sysbio/syad010
Mueller B.D., Huttenlocker A.K., Small B.J., Pinto J.L., Dean-Wallace K., Chatterjee S. 2023. A new kannemeyeriiform dicynodont (Synapsida) from a Late Triassic vertebrate assemblage in West Texas, USA. J. Vertebr. Paleontol. 43(2):e2255236. https://doi.org/10.1080/02724634.2023.2255236

Mulvey L.P., May M.R., Brown J.M., Höhna S., Wright A.M., Warnock R.C. 2024. Assessing the adequacy of morphological models using posterior predictive simulations. Syst. Biol. syae055.
Napoli J. 2024. Resolving the "ontogeny problem" in vertebrate paleontology. bioRxiv. https://doi.org/10.1101/2024.10.25.620216
Nylander J.A., Ronquist F., Huelsenbeck J.P., Nieves-Aldrey J. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53(1):47–67. https://doi.org/10.1080/10635150490264699
Olivier C., Battail B., Bourquin S., Rossignol C., Steyer J.S., Jalil N.E. 2019. New dicynodonts (Therapsida, Anomodontia) from near the Permo-Triassic boundary of Laos: implications for dicynodont survivorship across the Permo-Triassic mass extinction and the paleobiogeography of Southeast Asian blocks. J. Vertebr. Paleontol. 39(2):e1584745. https://doi.org/10.1080/02724634.2019.1584745
O'Reilly J.E., Puttick M.N., Parry L., Tanner A.R., Tarver J.E., Fleming J., Pisani D., Donoghue P.C. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. Biol. Lett. 12(4):20160081. https://doi.org/10.1098/rsbl.2016.0081
Parins-Fukuchi C. 2018. Use of continuous traits can improve morphological phylogenetics. Syst. Biol. 67(2):328–339. https://doi.org/10.1093/sysbio/syx072
Peters S.E., McClennen M. 2016. The paleobiology database application programming interface. Paleobiology 42(1):1–7. https://doi.org/10.1017/pab.2015.39
Pogue M.G., Mickevich M. 1990. Character definitions and character state delineation: the bête noire of phylogenetic inference. Cladistics 6(4):319–361. https://doi.org/10.1111/j.1096-0031.1990.tb00549.x
Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67(5):901–904. https://doi.org/10.1093/sysbio/syy032
Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53(1–2):131–147. https://doi.org/10.1016/0025-5564(81)90043-2
Schlager S., Profico A., Di Vincenzo F., Manzi G. 2018. Retrodeformation of fossil specimens based on 3D bilateral semi-landmarks: implementation in the R package "morpho". PloS One 13(3):e0194073. https://doi.org/10.1371/journal.pone.0194073.
Scotland R.W., Olmstead R.G., Bennett J.R. 2003. Phylogeny reconstruction: the role of morphology. Syst. Biol. 52(4):539–548. https://doi.org/10.1080/10635150309309
Simões T.R., Vernygora O.V., De Medeiros B.A., Wright A.M. 2023. Handling logical character dependency in phylogenetic inference: extensive performance testing of assumptions and solutions using simulated and empirical data. Syst. Biol. 72(3):662–680. https://doi.org/10.1093/sysbio/syad006
Smith R., Rubidge B., van der Walt M. 2013. Therapsid biodiversity patterns and paleoenvironments of the Karoo Basin, South Africa. In: Chinsamy-Turan A., editor. Forerunners of mammals: radiation, histology, biology. Bloomington: Indiana University Press. p. 31–62.
Smith R.M.H., Rubidge B.S., Day M.O., Botha J. 2020. Introduction to the tetrapod biozonation of the Karoo Supergroup. S. Afr. J. Geol. 123(2):131–140. https://doi.org/10.25131/sajg.123.0009
Sulej T., Niedźwiedzki G. 2019. An elephant-sized Late Triassic synapsid with erect limbs. Science 363(6422):78–80. https://doi.org/10.1126/science.aal4853
Sullivan C., Reisz R.R., Smith R.M. 2003. The Permian mammal-like herbivore *Diictodon*, the oldest known example of sexually dimorphic armament. Proc. R. Soc. London B 270(1511):173–178. https://doi.org/10.1098/rspb.2002.2189
Thiele K. 1993. The holy grail of the perfect character: the cladistic treatment of morphometric data. Cladistics 9(3):275–304. https://doi.org/10.1111/j.1096-0031.1993.tb00226.x
Thorne J., Kishino H., Painter I.S. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657. https://doi.org/10.1093/oxfordjournals.molbev.a025892
Tribble C.M., Freyman W.A., Landis M.J., Lim J.Y., Barido-Sottani J., Kopperud B.T., Höhna S., May M.R. 2022. RevGadgets: an R package for visualizing Bayesian phylogenetic analyses from RevBayes. Methods Ecol. Evol. 13(2):314–323. https://doi.org/10.1111/2041-210X.13750

Uhen M.D., Allen B., Behboudi N., Clapham M.E., Dunne E., Hendy A., Holroyd P.A., Hopkins M., Mannion P., Novack-Gottshall P., Pimiento C., Wagner P. 2023. Paleobiology database user guide version 1.0. PaleoBios 40(11). https://doi.org/10.5070/P9401160531

Viglietti P.A., Benson R.B., Smith R.M., Botha J., Kammerer C.F., Skosan Z., Butler E., Crean A., Eloff B., Kaal S. et al. 2021. Evidence from South Africa for a protracted end-Permian extinction on land. Proc. Natl. Acad. Sci. USA 118(17):e2017045118. https://doi.org/10.1073/pnas.2017045118.

Viglietti P.A., Smith R.M., Angielczyk K.D., Kammerer C.F., Fröbisch J., Rubidge B.S. 2016. The Daptocephalus Assemblage Zone (Lopingian), South Africa: a proposed biostratigraphy based on a new compilation of stratigraphic ranges. J. Afr. Earth Sci. 113: 153–164. https://doi.org/10.1016/j.jafrearsci.2015.10.011

Warnock R.C., Wright A.M. 2020. Understanding the tripartite approach to bayesian divergence time estimation. Cambridge University Press.

Warren D., Geneva A., Swofford D., Lanfear R. 2016. rwty: R we there yet. A package for visualizing MCMC convergence in phylogenetics.

Webster M., Hughes N.C. 1999. Compaction-related deformation in Cambrian olenelloid trilobites and its implications for fossil morphometry. J. Paleontol. 73(2):355–371. https://doi.org/10.1017/S0022336000027827

Whidden C., Matsen IV F.A. 2015. Quantifying MCMC exploration of phylogenetic tree space. Syst. Biol. 64(3):472–491. https://doi.org/10.1093/sysbio/syv006

Wickham H. 2011. ggplot2. Wiley Interdiscip. Rev. Comput. Stat. 3(2):180–185. https://doi.org/10.1002/wics.147

Wright A.M. 2019. A systematist's guide to estimating Bayesian phylogenies from morphological data. Insect Syst. Div. 3(3):2. https://doi.org/10.1093/isd/ixz006

Wright A.M., Hillis D.M. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. PloS One 9(10):e109210. https://doi.org/10.1371/journal.pone.0109210

Wright D.F., Hopkins M.J. 2025. Assessing the impact of character evolution models on phylogenetic and macroevolutionary inferences from fossil data. bioRxiv. https://doi.org/10.1101/2024.12.23.630137

Wright A.M., Lloyd G.T., Hillis D.M. 2016. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. Syst. Biol. 65(4):602–611. https://doi.org/10.1093/sysbio/syv122

Wright A.M., Wynd B.M. 2024. Modeling of rate heterogeneity in datasets compiled for use with parsimony. bioRxiv. https://doi.org/10.1101/2024.06.26.600858

Wynd B.M., Peecook B.R., Whitney M.R., Sidor C.A. 2017. The first occurrence of *Cynognathus crateronotus* (Cynodontia: Cynognathia) in Tanzania and Zambia, with implications for the age and biostratigraphic correlation of Triassic strata in southern Pangea. J. Vertebr. Paleontol. 37(sup1):228–239. https://doi.org/10.1080/02724634.2017.1421548

Wynd B.M., Uyeda J.C., Nesbitt S.J. 2021. Including distorted specimens in allometric studies: linear mixed models account for deformation. Integrative Org. Biol. 3(1):obab017. https://doi.org/10.1093/iob/obab017

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39(3):306–314. https://doi.org/10.1007/BF00160154

Zhang R., Drummond A.J., Mendes F.K. 2024. Fast Bayesian inference of phylogenies from multiple continuous characters. Syst. Biol. 73(1):102–124. https://doi.org/10.1093/sysbio/syad067

Syst. Biol. 00(0):1–17, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of the Society of Systematic Biologists. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site-for further information please contact journals.permissions@oup.com
https://doi.org/10.1093/sysbio/syaf078
Advance Access publication 0, 2025