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, Louisiana, USA*

[2] *GeoBio-Center, Ludwig-Maximilians-Universität München, 80333 Munich, Germany*

[3] *Department of Earth and Environmental Sciences, Paleontology & Geobiology, Ludwig-Maximilians-Universität München, 80333 Munich, Germany*

[4] *North Carolina Museum of Natural Sciences, Raleigh, North Carolina, USA*

[5] *Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina, USA*

[6] *Dept. of Earth Atmospheric Sciences, and School of Biological Sciences, University of Nebraska, Lincoln, Nebraska, USA*

[*] *Corresponding Author: april.wright@selu.edu*

March 3, 2025

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 manuscript, 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 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.

Continuous traits, Discrete traits, Phylogenetic methods, Morphological data, Bayesian phylogenetics, RevBayes

1 Introduction

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, 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 (Scotland et al., 2003; Brazeau, 2011), 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. That is, 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). The nearest exception concerns presence/absence characters, represented by either a 0 or 1 for gain or loss of a given trait. These are ubiquitous: across sampled morphological datasets, roughly 70% of characters are binary in this manner (Barido-Sottani et al., 2020). However, even here the comparison to coding insertions/deletions in molecular matrices is inexact: although "absent" is (on one level) objectively comparable, "present" often represents different and not necessarily equivalent additions to the organisms' morphology.

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). Historically, parsimony has been the dominant phylogenetic tool for paleontologists. With the growing usage of Bayesian methods in morphological phylogeny estimation, most methods have been evaluated for discrete characters (Wright and Hillis, 2014; O'Reilly et al., 2016; Wright et al., 2016; Klopfstein et al., 2015, 2019) because of their ready availability from prior datasets intended for parsimony-based analyses. However, adapting these data for Bayesian analyses incorporates biases associated with how to discretize each character, and means that 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.

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 (though ratios tend to be favored, as raw measurements are more likely to be influenced by ontogeny when a sample includes differently sized individuals; Bardin et al. (2017); Napoli (2024); Griffin et al. (2021)). Direct measurements of anatomical features, such as lengths or angulations of structures or

ratios of one structure to another, have the potential to be much 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), while others argue that they are the "Holy Grail" of phylogenetic characters (Thiele, 1993). 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 datasets, 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 minimums-steps parsimony methods to infer phylogeny, and thus might not pertain to likelihood and Bayesian analyses. Parins-Fukuchi (2018) used a Brownian Motion (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. Brownian Motion 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.

Parsimony-based cladistics has been 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). Although many methods exist to model the evolution of continuous traits on a tree, few have explored how these methods might behave in actually estimating the tree (Parins-Fukuchi, 2018). Here, we explore the utility of using character evolution models of Brownian Motion 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). 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 provides the flexibility to combine models according to the data in use, though see May and Rothfels (2023). 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.

For this, we rely on a recent phylogeny of Dicynodontia, a group of early synapsids (the mammalian stem lineage) originating during the Permian Period. Dicynodonts were a diverse lineage of mammal relatives that lived from the middle Permian through the Late Triassic periods (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). We chose to work with this group 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 the dicynodont phylogeny (Kammerer et al., 2019). 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. With these data, we explore the utility of Brownian Motion character evolution models in the estimation of phylogenetic trees in the software RevBayes (Höhna et al., 2016). We explore phylogenies estimated by continuous character evolution models both independently of and co-estimated with discrete characters. We performed this work in RevBayes to provide a user-friendly option for total-evidence tip-dated phylogenetic estimation with both continuous and discrete models of character evolution.

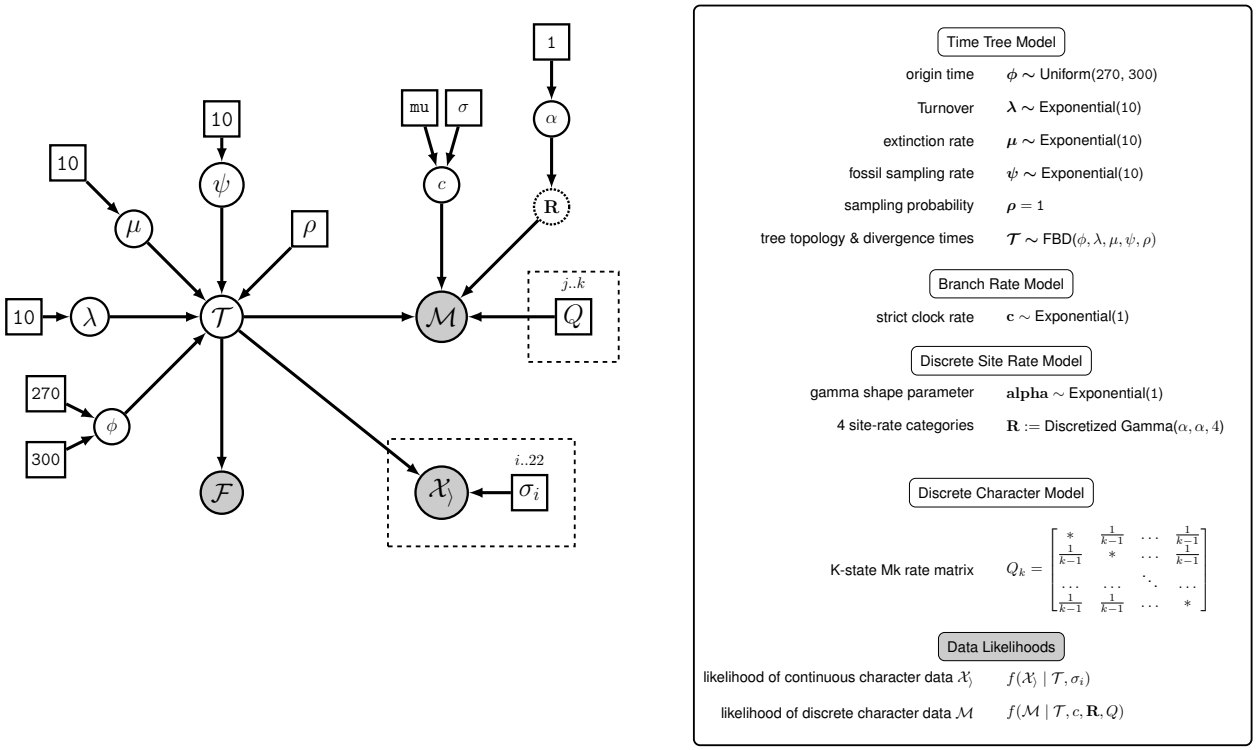

Figure 1: A graphical model displaying the components of the full joint phylogenetic model. An explanation of the model components is displayed at right. Adapted from Barido-Sottani et al. (2020).

2 Materials and Methods

2.1 Background

The inclusion of continuous character information in parsimony-based frameworks treats them as additive, such that 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 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 Brownian Motion 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 Brownian motion and additive character evolution models are quite similar, they are computationally distinct, as Brownian motion 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 Brownian motion would likely be more suitable in a Bayesian framework, wherein the estimation of branch lengths is essential to the estimation of the tree.

Throughout this manuscript we are testing the utility of including both continuous and discrete morphological character information in joint (also called total-evidence) phylogenetic inference. The inference model can then be further subdivided in three constituent parts: the substitution, clock, and tree models.

The substitution model establishes the evolutionary relationships between different taxa (Fig. 1). For the molecular data, there has been development of several substitution models that explicitly follow expected transitions between nucleotide types (e.g., JC69 (Jukes and Cantor, 1969), F81 (Felsenstein, 1981), HKY (Hasegawa et al., 1985), GTR (Tavaré, 1986)). Because discrete morphological characters lack the same degree of objectivity as in molecular characters, the established substitution models are unsuitable other than JC69 (Lewis, 2001; Huelsenbeck et al., 2004; Khakurel et al., 2024). To ease calculations and allow for more appropriate comparisons between inference methods (e.g., parsimony vs Bayesian), morphological characters have largely been confined to using the Mk model (Lewis, 2001) for discrete character evolution. This model is the morphological counterpart of the JC69 model which assumes that the rate of change between the character states are equal. Additionally, although parsimony does not contain any information about rates per se, equal-weights parsimony assumes that all characters have an equivalent contribution to the final tree score (steps), which is effectively what the Mk model is doing in likelihood-based inferences. There have been several variations of the Mk model which allow researchers to relax the assumptions of the rates of changes (Nylander et al., 2004; Klopfstein et al., 2015; Wright et al., 2016), but are not used here, as mismatch of the morphological substitution process is mostly unproblematic for dated analyses (Klopfstein et al., 2019).

Continuous morphological characters have received comparatively little attention in regards to Bayesian phylogenetic inference (but see Parins-Fukuchi (2018) and Zhang et al. (2024)), but have been modeled extensively onto existing phylogenies using phylogenetic comparative methods. Phylogenetic comparative methods take advantage of time-calibrated phylogenies, alongside continuous and discrete variables, to model the tempo and mode of evolution between species. The evolution of continuous variables under these comparative methods has primarily been modeled under Brownian motion (i.e., random-walk: (Felsenstein, 1973, 1985a; Gingerich, 1993)) and Ornstein-Uhlenbeck (i.e., directional evolution: (Beaulieu et al., 2012; Hansen, 1997; Butler and King, 2004)), with many variants existing to relax or constrain model assumptions. These models allow for changes to accumulate continuously along branches, and because Brownian motion character evolution models conceptually mirror the additive method, character evolution models should be of similar utility in Bayesian inference as in parsimony.

The second component is the clock model, and this portion describes the variation of the evolutionary rates across the tree (Fig. 1). There can be several types of clock models, including strict clock (Zuckerkandl and Pauling, 1962), uncorrelated clock (Drummond et al., 2006), and auto-correlated clock (Kishino et al., 2001). The third portion is the tree model, which describes the process of speciation, extinction, and lineage sampling that leads to the observed tree. For the purposes of this study, these two components will be constant across all estimations. Herein, we are explicitly testing the substitution model component, and evaluating the utility of continuous character evolution models in Bayesian phylogenetic inference while holding the other two components constant.

2.2 Dicynodontia as a Study System

Dicynodontia is a clade of Permian and Triassic synapsids that represents an ideal test case to evaluate the impact of continuous morphological characters on discrete morphological character matrices. 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; Wynd et al., 2017; Viglietti et al., 2016). Dicynodonts are represented by 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., 2020b). 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). The later Triassic kannemeyeriiforms (particularly the Stahleckeriidae) tend to be known from less complete materials 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 from single damaged skeletal elements (e.g.,

Eubrachiosaurus browni), or extensive but eroded and disarticulated remains (e.g., *Placerias hesternus*) (Kammerer et al., 2013; Fiorillo et al., 2000). Observed differences in postcranial anatomy between these taxa are largely proportional, and dicynodont workers have attempted to capture this variation through 23 continuous character measurements (primarily ratios). When incorporated under an additive model alongside morphological character data into parsimony-based phylogenetics, these characters have been demonstrated to provide additional phylogenetic resolution unavailable using only discrete-state characters (Kammerer et al., 2011). All analyses of dicynodont phylogeny from the past decade are based on the underlying dataset of Kammerer et al. (2011), which introduced continuous character data to this system. Because continuous measurements have been used, and used consistently, in the works that established the currently accepted dicynodont tree, this represents an ideal system to test the utility of using character evolution models alongside morphological characters in a Bayesian framework. Not only will this 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 them to behave.

2.3 Character Evolution Models

2.3.1 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 are 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:

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

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 dataset. For the Mk model, we established among character rate variation by using gamma distributed rate variation, with four discrete rate categories (Yang, 1994).

2.3.2 Brownian Motion

The Brownian motion model (Felsenstein, 1973; Parins-Fukuchi, 2018), 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.

As a model driven purely by drift, the Brownian motion 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 Brownian Motion model:

- Strict Brownian Motion: Under this model, all characters share a *sigma* value drawn from a uniform distribution bounded between 1e-5 and 1e-1. This model will be referred to as **Continuous Strict** in figures and further text.

- Relaxed Brownian Motion: Under this model, each character has an independent *sigma*, drawn from the uniform distribution described above. This model will be referred to as **Continuous Relaxed** in figures and further text.

2.3.3 Fossilized Birth-Death Model

For the Bayesian inference, the model requires starting values/distributions for each of the parameters for the model to pull from during the climb toward convergence, and are known as the priors. 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 archived 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 expectations 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 million years old. We set the outgroup to include *Biarmosuchus tener* and *Hipposaurus boonstrai*, which share a most recent common ancestor and belong to the clade Biarmosuchia (Sidor et al., 2021). No clade constraints were used in our analyses when comparing the continuous, discrete, and joint models to one another. Rather, starting trees where drawn from a distribution of trees plausible under the FBD distribution. We explored whether sampled ancestors are present or not, but decided not to include them here as we were consistently oversampling the number of ancestors. For the fossil ages, we used occurrences from the Paleobiology Database (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* zone) is thought to span from 255.2 - 254.0 Ma (Smith et al. 2020a). Simply assuming that *Dicynodon* occurrences fall within that span would present circular reasoning. However, the zone taxa analyzed here appear *before* their eponymous zones, indicating that the other criteria we use to date PBDB sites 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: http://datadryad.org/stash/share/ahyNsjmSGsFvdaQ0SvVhmUj2JmpOhnhnlK8Wlp_ensQ.

2.3.4 Clock Models

We used an uncorrelated lognormal clock to model the rate of evolution across the tree (Thorne et al., 1998). This clock model is very flexible, allowing ancestor and descendant branches to have very different rates of evolution.

2.3.5 Joint Inference

We ran a total of four different models that include varying amounts of discrete and continuous character information (Table 1): continuous with all characters sharing a Brownian Motion rate (**continuous-strict**), continuous with characters having their own BM rate (**continuous - relaxed**), discrete characters under the Mk model and continuous characters in which all continuous characters share a BM rate(**combined - strict**), and discrete and continuous characters with continuous characters having their own BM rate (**combined - relaxed**), all of which were compared to the discrete-only model.

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
Discrete	Yes	No	NA
Continuous - Strict	No	Yes	1
Continuous - Relaxed	No	Yes	23
Combined - Strict	Yes	Yes	1
Combined - Relaxed	Yes	Yes	23

2.4 Bayesian inference

We performed phylogenetic inferences in `RevBayes`, 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). We set a 25 percent burn-in to our sample, which effectively removes the portion of the MCMC that is climbing toward convergence. 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. The practical difference between Bayesian or parsimony based inference is that parsimony seeks to produce the fewest number of trees that minimize the total number of character transitions required, whereas Bayesian inference generates a large number of plausible topologies (often thousands to hundreds of thousands of trees) and records their parameter estimates. Bayesian inference seeks to account for the uncertainty that is ubiquitous to life by approximating a reasonable solution and iterating on that solution repeatedly to fully explore the uncertainty that surrounds our most likely topologies.

2.5 Assessing model convergence

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). Estimated sample size is a correction to account for the autocorrelation present in the posterior sample, and a parameter having an ESS under 200 is generally considered not to be converged. Posterior samples were checked in Tracer v1.7.2, and tracer plots for individual parameters (e.g., likelihood) are checked by eye for a general slope greater than 1.0, which implies that the model was still 'climbing' and had yet to reach convergence (Rambaut et al., 2018).

Table 2: Number of `RevBayes` generations needed for the model to converge.

Name	Generations to Convergence
Discrete	260,000
Continuous - Strict	235,000
Continuous - Relaxed	135,000
Combined - Strict	200,000
Combined - Relaxed	200,000

2.6 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.

3 Results

3.1 Model results

All model and data combinations ultimately reach convergence in our experiments. Plots on Fig. 4 are shown downsampled, as to appear on the same axis. However, on Table 2, the number of generations required for convergence is shown. Interestingly, the discrete-only datasets take the longest to converge, followed by the combined - strict analysis. The combined-relaxed 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 to other software packages that can use continuous traits, such as `BEAST 2` (Bouckaert et al., 2014).

The different treatments arrive at different origin times for the clade. Most treatments favor an origin time of between 280-270 million years before the present, which is consistent with the estimated age of divergence between Biarmosuchia, Dinocephalia, Anomodontia, and Theriodontia based on recent analyses examining therapsid interrelationships more broadly (Matamales-Andreu et al., 2024). However, the combined-relaxed dataset and model combination prefers a much older origin time, prior to 300Ma. This is unlikely, 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 ∼270 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. Additionally, the origin times for the combined-relaxed analysis converge around 320-310 MA, which would predate the earliest known synapsids (Mann et al., 2023) and indicate that the observed signal in the fossil record is completely erroneous.

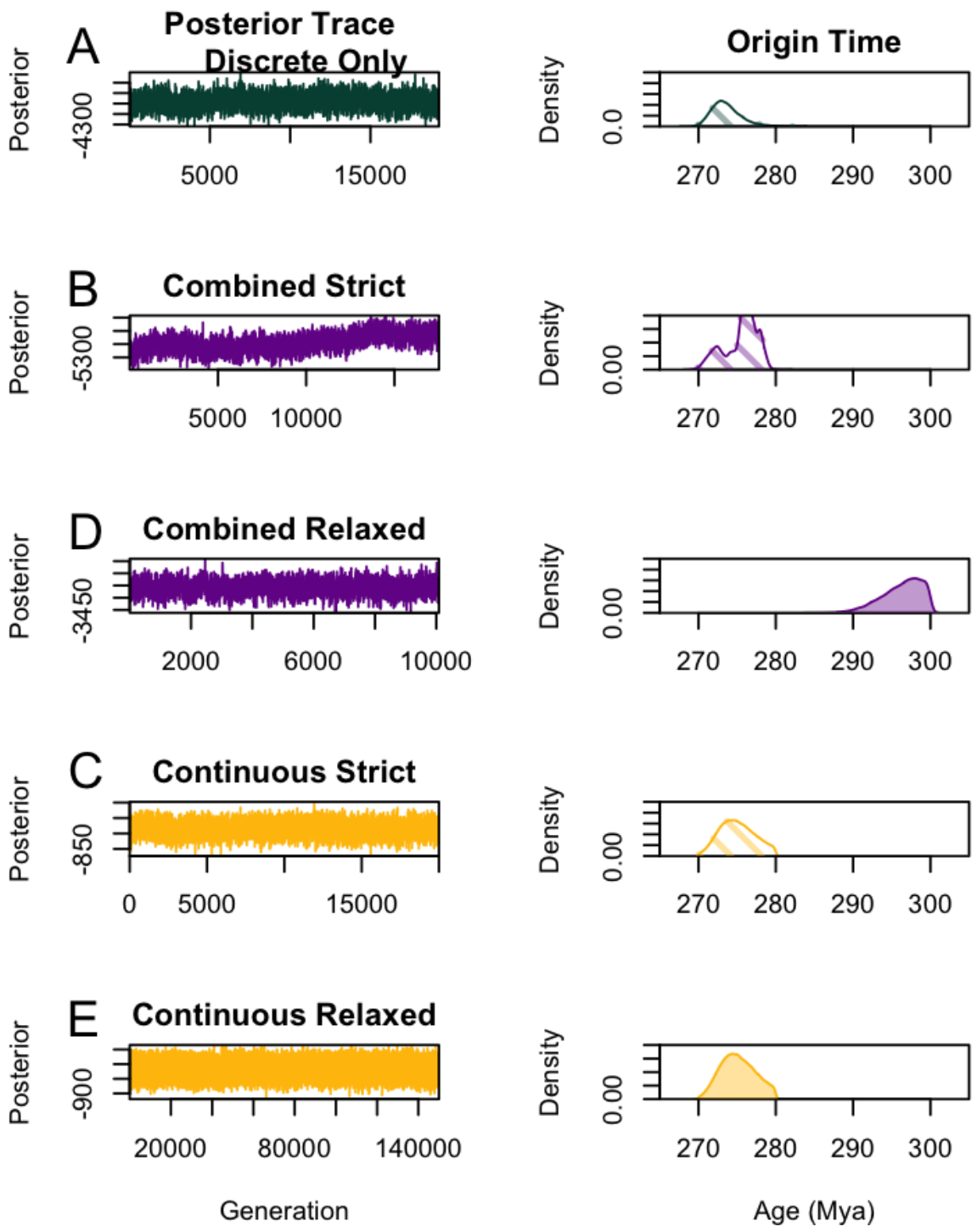

Figure 2: **Visualization of the tracer plots for each analysis. Colloquially referred to as the "Fuzzy Caterpillar", convergence is assumed when the caterpillar is horizontal and sampling the same distribution. In order to fit all the traces on the same X-axis, the number of generations is down sampled by a factor of 100. At right, the origin times favored in different analyses is shown.**

3.2 Tree topology

Our analyses did not reproduce the consensus tree that has been produced under cladistic analysis. We find that in general, when continuous variables are examined alone, the produced trees are essentially nonsense, producing erroneous relationships, and dissolving expected clades. When the discrete data are examined alone, we find that Kannemeyeriiformes is mostly intact, but a few kannemeyeriiform taxa are instead reconstructed as being a sister taxon to Permian dicynodonts traditionally recovered as distant to this group. Independently, the impact of fossilization is evident in the characters, where expected relationships amongst 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.

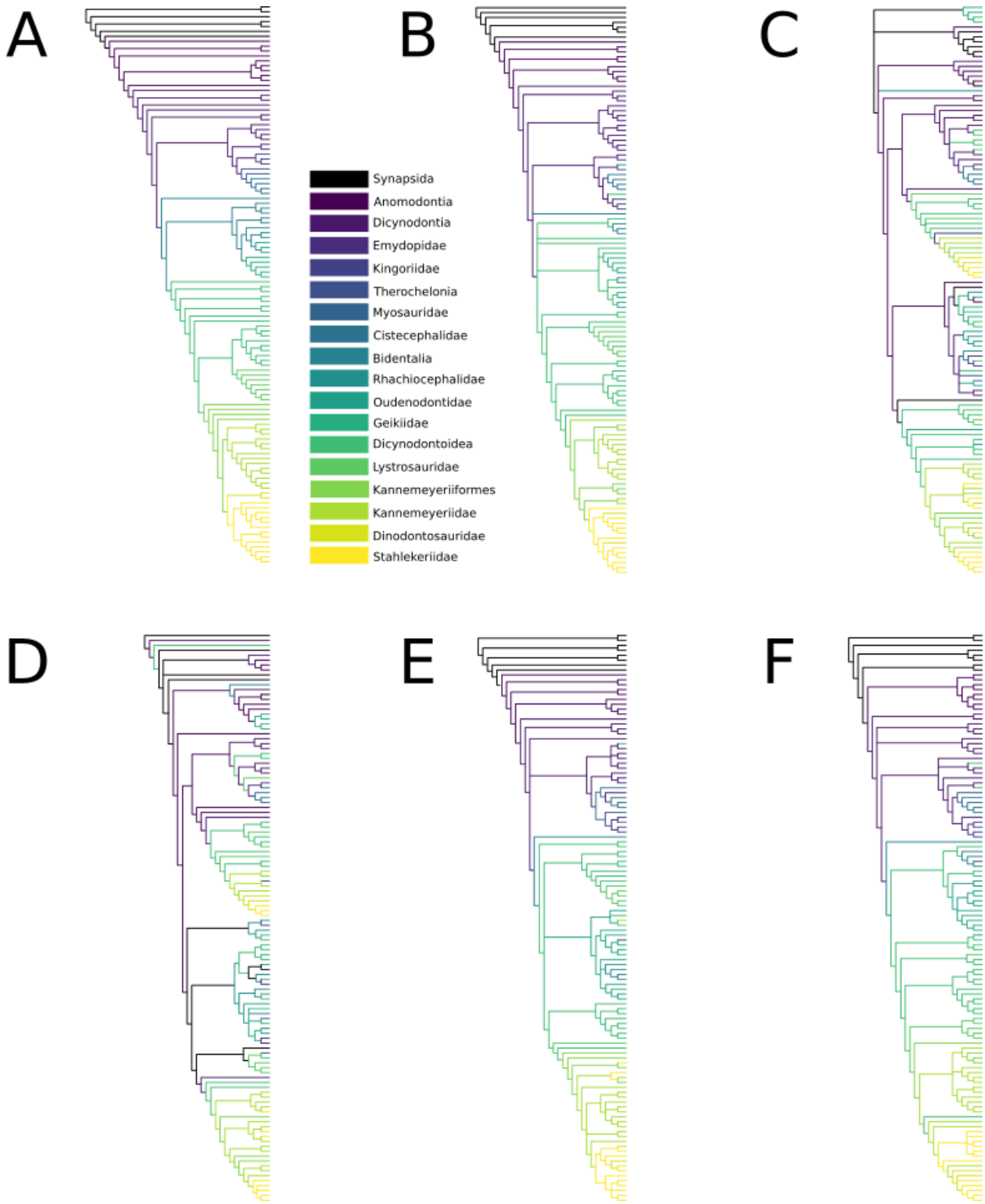

Figure 3: **Topological differences between models plot depicting how the monophyletic clades are differently dissolved in our various analyses. A) the currently accepted topology. B-F are maximum clade credibility trees (MCC) from our five models: B) Discrete only; C) Continuous Strict; D) Continuous Relaxed; E) Combined Strict; F) Combined Relaxed. 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 S1-S5.**

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. *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 to taxa known from much more fragmentary material. The case for *Dicynododontoides* should be even more concerning, as this taxon, typically a non-bidentalian emydopoid, is recovered here as belonging to Bidentalia (Discrete-only and Combined relaxed), among lystrosaurids (Continuous only-strict and relaxed), and as sister taxon to the Geikiidae (Combined strict). *Dicynodontoides* 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). Some differences in topology relative to the 'standard' parsimony-based trees are expected for our analyses; for example, while 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' positions of exceptionally well-represented taxa such as *Ulemica* and *Dicynodontoides*, with extremely well-supported placements is previous trees, is highly concerning, and suggests that the explanatory power of synapomorphies in parsimony analyses do not readily transfer to Bayesian total evidence analyses.

3.3 Treespace

When we generate a treespace comparing discrete characters, continuous characters, and a combination of the two, we see that the combined analysis plots primarily with the continuous-only data. The discrete data occupy a separate part of the treespace, with a few trees approaching the treespace shared by the other analyses. This indicates a disproportionate impact of continuous characters on combined analyses when compared with their discrete counterparts.

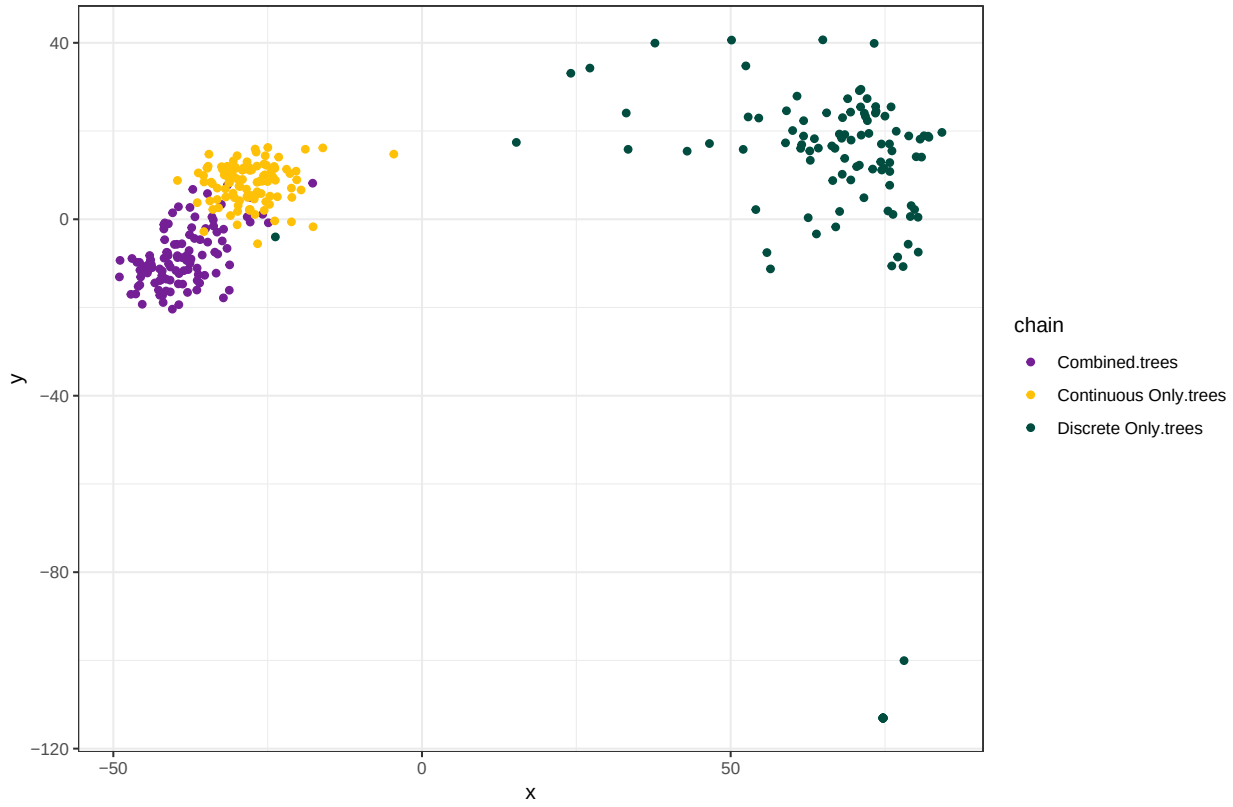

Figure 4: **A treespace depicting the trees estimated without fossil age information using the tree different data types. The metric is Robinson Foulds. Thus, points that are closer together are more similar topologically and in branch lengths.**

3.4 Tree lengths

To gain a sense of how model choice impacts phylogenetic inference, we estimate treelength, as a cumulative sum of all internal branches within the tree. The discrete, continuous (relaxed BM), and combined (strict BM) models produce similar tree lengths. Continuous characters modeled under a strict clock favor longer trees, but the longest trees come from jointly modeling discrete and continuous characters with a relaxed BM model. This appears mostly driven by this dataset favoring an older origin time for the total group (Fig. 2).

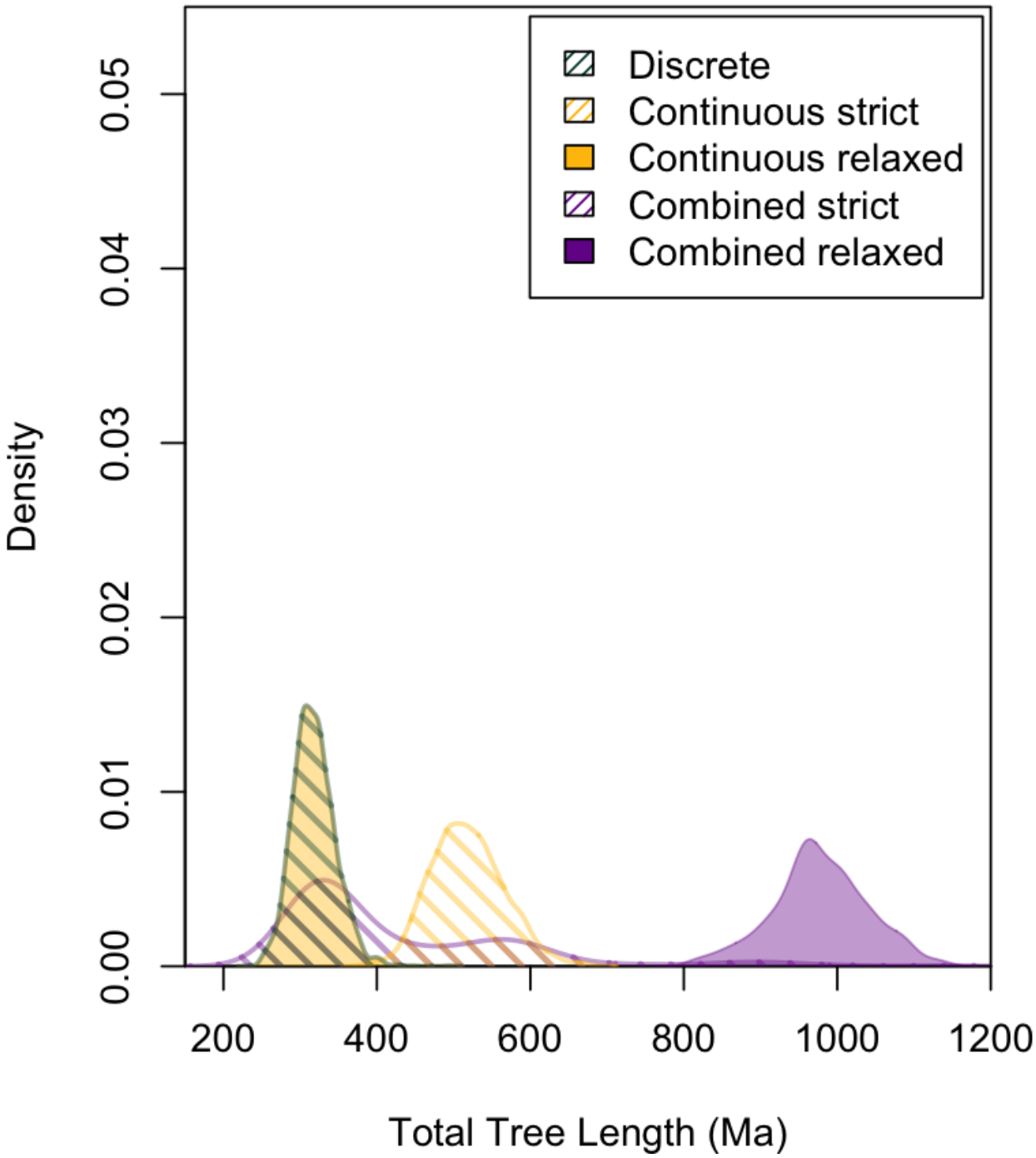

Figure 5: Tree lengths in millions of years for the different estimations.

4 Discussion

4.1 The unintended consequences of continuous characters

Continuous variables preserve morphological characteristics without introducing subjectivity. A core issue, however, is that direct quantification of morphology tends to either obscure nuance when changes are textural, or they require high-level 3D scans accompanied by advanced computational methods that can produce and compare topological maps of morphology (Goswami et al., 2014; Fabre et al., 2020; Boyer, 2008). 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 appear to have a large effect on topological results and the manner in which treespace is explored. 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). 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 datasets (Bapst et al., 2018). This is particularly problematic, as the continuous variables are largely used to better place the kannemeyeriiform dicynodonts and their generally disarticulated preservation when compared to Permian dicynodonts, which are found to be interspersed throughout the tree when exploring the discrete-only model. 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 complications surrounding continuous variables in phylogenetics, especially in the context of branch lengths, is potentially due to model misspecification. When examining our MCMC parameters, we find what appear to be opposing results when continuous characters are treated under a single rate or individual rates for each continuous character, either alone or with discrete characters. Although these results seem counterintuitive, they make sense when discordance is considered. Attempting to generate a tree only using continuous characters is far easier when multiple different rates are allowed to exist. 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. On the other hand, when we incorporate discrete characters, it is simpler to include the continuous characters with a strict BM. This is likely due to the simplicity when the MCMC only has to adjust a single rate parameter once between each generation. When the relaxed clock is implemented, 23 separate parameters need to be adjusted each iteration, increasing model complexity and computation time. However, despite this complexity, the model achieves good convergence fairly quickly.

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., Brownian motion), 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., Ornstein-Uhlenbeck 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. Brownian motion 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, however are not currently implemented for inference to test.

Joint analyses of discrete and continuous traits still produce interesting artifacts. As seen on Fig. 5, continuous traits combined with discrete traits and a relaxed Brownian Motion processes produce much longer tree lengths either discrete or continuous traits alone. Conceptually, this makes sense as to why it would be a problem in the Bayesian analysis, but not the cladistic analysis. 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. In the case of Brownian motion, 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.

4.2 The benefits of including continuous variables in joint phylogenetic estimation

The issues pointed out above really point to the utility of continuous characters and how we need to thoroughly vet both our data and how it behaves under differing model circumstances. As seen on Fig. 4, continuous traits can have a powerful impact on the treespace explored. Importantly, when more than a single rate is specified for the Brownian model, there is no major change to the origin time, and branch lengths are consistent with those generated through discrete characters under an Mk model (Figs. 2, 5). We discuss above how, aside from branch lengths, the additive model is conceptually similar to a character evolution model following Brownian Motion; however, the additive model lacks a unifying rate parameter, and because all calculations are done at nodes, the additive model can effectively incorporate characters with lots of rate heterogeneity and not have it majorly impact the results. The 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.

When it comes to dicynodont topology, the discrete and combined-strict analyses produce topologies most consistent with the currently-supported dicynodont relationships. More specifically, the placement of the Permian dicynodonts follow overall expectations (the more complete specimens of the sample), with the reconstruction of Triassic Kannemeyeriiform relationships being more spurious. Because the entire point behind the use of continuous characters in this analysis is to better place the Kannemeyeriiformes, our 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 the Kannemeyeriiforms.

A similar story is found in the divergence dates, where the discrete, continuous strict, and continuous relaxed analyses produce clade ages more or less in line with reality, and the combined relaxed analyses producing much older clade ages than expected. The more striking concerns in the taxon ages are exemplified in the temporal disconnect between the two species of *Eodicynodon*. Although both *"Eodicynodon" oelofseni* and *Eodicynodon oosthuizeni* were found in the same quarry, mere meters away from one another, only the combined-relaxed 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.

The incorporation of multiple sources of data is nearly guaranteed to result in increased model complexity and trade-offs between the different forms of data. It may come as no surprise that there is a trade-off that has been limiting the use of continuous characters, and that is computational time and resources. Creating

a single rate parameter and applying it across all continuous characters will generate results rapidly, but will most likely fail to form the expected clades. On the other hand, modeling a unique rate parameter for each character, and allowing that parameter to vary between iterations in the MCMC is going to increase computational time, and could limit many researchers to requiring university supercomputers to perform their analyses. The use of superclusters, interconnected computer servers that allow for long-term computational projects to run uninterrupted on state of the art machinery, is not uncommon in the world of molecular phylogenetics, but it is more uncommon in the world of morphological and paleontological phylogenetics. More research is needed into the efficiency of methods to encourage widespread adoption.

4.3 An objective future for fossil phylogenetics

Molecular phylogenetics has long established incorporating multiple forms of information for joint estimation, often labeled as 'total evidence' analysis. Simultaneously continuous characters have kneecapped investigation because of computational limits surrounding infinite traitspace. However, methods surrounding the application of continuous variables modeled on existing dated trees, often described as the comparative method, have established numerous character evolution models, including Brownian Motion 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 continuous variables explore distinct regions of traitspace that may be otherwise uninvestigated. Uniting the simplest character evolution model (BM) along with the simplest morphological model (Mk) establishes a critical starting point for advancing phylogenetic inference in the life sciences.

Likelihood-based phylogenetics has historically attempted to incorporate continuous variables through threshold models, which discretize continuous variables that have numerical values exceeding a threshold (Felsenstein, 1988). This represents some of the earliest work in incorporating character evolution models into the estimation of a phylogenetic tree (but see Parins-Fukuchi (2018)). 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 misspecification 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 Ornstein-Uhlenbeck model which is meant to be an approximation of selection, will be necessary to obtain our best approximations of phylogeny.

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 to be a future paradigm shift in the field. Joint inference integrates neontological concepts and tools, such as clock models, with classic 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.

5 Acknowledgments

AMW, BMW, and BK were supported on NSF DEB 2045842 and AMW was supported on NSF CIBR 2113425. AMW additionally received funds from Southeastern Louisiana University's Dyson Fellowship. PJW was supported on NSF EAR 2129628. BK was additionally supported by the European Union (ERC, MacDrive, GA 101043187). 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.

6 Data Availability:

All the original data and scripts necessary to reproduce the analyses reported in this study can be accessed through the Dryad link: http://datadryad.org/stash/share/ahyNsjmSGsFvdaQ0SvVhmUj2Jmp0hnhnlK8Wlp_ensQ

References

Angielczyk, K. D., P. J. Hancox, and A. Nabavizadeh. 2017. A redescription of the triassic kannemeyeriiform dicynodont sangusaurus (therapsida, anomodontia), with an analysis of its feeding system. Journal of Vertebrate Paleontology 37:189–227.

Angielczyk, K. D., C. A. Sidor, S. J. Nesbitt, R. M. Smith, and L. A. Tsuji. 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). Journal of Vertebrate Paleontology 29:1174–1187.

Bapst, D. W., H. A. Schreiber, and S. J. Carlson. 2018. Combined analysis of extant rhynchonellida (brachiopoda) using morphological and molecular data. Systematic Biology 67:32–48.

Bardin, J., I. Rouget, and F. Cecca. 2017. Ontogenetic data analyzed as such in phylogenies. Systematic biology 66:23–37.

Barido-Sottani, J., J. A. Justison, A. M. Wright, R. C. Warnock, W. Pett, and T. A. Heath. 2020. Estimating a time-calibrated phylogeny of fossil and extant taxa using RevBayes. Pages 5.2:1–5.2:23 *in* Phylogenetics in the Genomic Era (C. Scornavacca, F. Delsuc, and N. Galtier, eds.). No commercial publisher — Authors open access book.

Barido-Sottani, J., O. Schwery, R. C. Warnock, C. Zhang, and A. M. Wright. 2024. Practical guidelines for bayesian phylogenetic inference using markov chain monte carlo (mcmc). Open Research Europe 3:204.

Beaulieu, J. M., D.-C. Jhwueng, C. Boettiger, and B. C. O'Meara. 2012. Modeling stabilizing selection: expanding the ornstein–uhlenbeck model of adaptive evolution. Evolution 66:2369–2383.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10:e1003537.

Boyer, D. M. 2008. Relief index of second mandibular molars is a correlate of diet among prosimian primates and other euarchontan mammals. Journal of human evolution 55:1118–1137.

Brazeau, M. D. 2011. Problematic character coding methods in morphology and their effects. Biological Journal of the Linnean Society 104:489–498.

Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. The American Naturalist 164:683–695.

Chappill, J. A. 1989. Quantitative characters in phylogenetic analysis. Cladistics 5:217–234.

Congreve, C. R., M. E. Patzkowsky, and P. J. Wagner. 2021. An early burst in brachiopod evolution corresponding with significant climatic shifts during the great ordovician biodiversification event. Proceedings of the Royal Society B: Biological Sciences 477:20211450.

Drummond, A., S. Ho, M. Phillips, and A. Rambaut. 2006. Relaxed Phylogenetics and Dating with Confidence. PLoS Biology 4:e88.

Fabre, A.-C., C. Bardua, M. Bon, J. Clavel, R. N. Felice, J. W. Streicher, J. Bonnel, E. L. Stanley, D. C. Blackburn, and A. Goswami. 2020. Metamorphosis shapes cranial diversity and rate of evolution in salamanders. Nature ecology & evolution 4:1129–1140.

Farris, J. S. 1970. Methods for computing wagner trees. Systematic Biology 19:83–92.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am J Hum Genet 25:471–92.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368–376.

Felsenstein, J. 1985a. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Felsenstein, J. 1985b. Phylogenies and the comparative method. The American Naturalist Pages 1–15.

Felsenstein, J. 1988. Phylogenies and quantitative characters. Annual Review of Ecology and Systematics Pages 445–471.

Fiorillo, A. R., K. Padian, and C. Musikasinthorn. 2000. Taphonomy and depositional setting of the placerias quarry (chinle formation: Late triassic, arizona). Palaios 15:373–386.

Gingerich, P. D. 1993. Quantification and comparison of evolutionary rates. American Journal of Science 293:453.

Goloboff, P. A., C. I. Mattoni, and A. S. Quinteros. 2006. Continuous characters analyzed as such. Cladistics 22:589–601.

Goswami, A., J. B. Smaers, C. Soligo, and P. D. Polly. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. Philosophical Transactions of the Royal Society B: Biological Sciences 369:20130254.

Griffin, C. T., M. R. Stocker, C. Colleary, C. M. Stefanic, E. J. Lessner, M. Riegler, K. Formoso, K. Koeller, and S. J. Nesbitt. 2021. Assessing ontogenetic maturity in extinct saurian reptiles. Biological Reviews 96:470–525.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22:160–174.

Höhna, S., T. A. Heath, B. Boussau, M. J. Landis, F. Ronquist, and J. P. Huelsenbeck. 2014. Probabilistic graphical model representation in phylogenetics. Systematic Biology 63:753–771.

Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Systematic Biology 65:726–736.

Huelsenbeck, J., B. Larget, and M. Alfaro. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. Molecular Biology and Evolution 21:1123.

Hunt, G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. Proceedings of the National Academy of Sciences 104:18404–18408.

Ivakhnenko, M. F. 1996. Primitive anomodonts, venyukoviids, from the late permian of eastern europe. Paleontological Journal 30:575–582.

Jardine, N. 1969. The observational and theoretical components of homology: a study based on the morphology of the dermal skull-roofs of rhipidistian fishes. Biological Journal of the Linnean Society 1:327–361.

Jukes, T. and C. Cantor. 1969. Evolution of protein molecules. Mammalian Protein Metabolism 3:21–132.

Kammerer, C. F., K. D. Angielczyk, and J. Fröbisch. 2011. A comprehensive taxonomic revision of dicynodon (therapsida, anomodontia) and its implications for dicynodont phylogeny, biogeography, and biostratigraphy. Journal of Vertebrate Paleontology 31:1–158.

Kammerer, C. F., S. Bandyopadhyay, and S. Ray. 2016. A new taxon of cistecephalid dicynodont from the upper permian kundaram formation of india. Papers in Palaeontology 2:569–584.

Kammerer, C. F., J. Fröbisch, and K. D. Angielczyk. 2013. On the validity and phylogenetic position of eubrachiosaurus browni, a kannemeyeriiform dicynodont (anomodontia) from triassic north america. PLoS One 8:e64203.

Kammerer, C. F. and M. d. l. A. Ordoñez. 2021. Dicynodonts (therapsida: Anomodontia) of south america. Journal of South American Earth Sciences 108:103171.

Kammerer, C. F., P. A. Viglietti, P. J. Hancox, R. J. Butler, and J. N. Choiniere. 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. Journal of Vertebrate Paleontology 39:e1596921.

Khakurel, B., C. Grigsby, T. D. Tran, J. Zariwala, S. Höhna, and A. M. Wright. 2024. The fundamental role of character coding in bayesian morphological phylogenetics. Systematic Biology 73:861–871.

Kishino, H., J. L. Thorne, and W. J. Bruno. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Molecular Biology and Evolution 18:352–361.

Klopfstein, S., R. Ryer, M. Coiro, and T. Spasojevic. 2019. Mismatch of the morphology model is mostly unproblematic in total-evidence dating: insights from an extensive simulation study. BioRxiv Page 679084.

Klopfstein, S., L. Vilhelmsen, and F. Ronquist. 2015. A nonstationary markov model detects directional evolution in hymenopteran morphology. Systematic biology 64:1089–1103.

Lanfear, R., X. Hua, and D. L. Warren. 2016. Estimating the effective sample size of tree topologies from bayesian phylogenetic analyses. Genome biology and evolution 8:2319–2332.

Lewis, P. O. 2001. A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. Systematic Biology 50:913–925.

Mann, A., A. C. Henrici, H.-D. Sues, and S. E. Pierce. 2023. A new carboniferous edaphosaurid and the origin of herbivory in mammal forerunners. Scientific Reports 13:4459.

Matamales-Andreu, R., C. F. Kammerer, K. D. Angielczyk, T. R. Simões, E. Mujal, Galobart, and J. Fortuny. 2024. Early-middle permian mediterranean gorgonopsian suggests an equatorial origin of therapsids. Nature Communications 15:1–12.

Matzke, N. J. and R. B. Irmis. 2018. Including autapomorphies is important for paleontological tip-dating with clocklike data, but not with non-clock data. PeerJ 6:e4553.

May, M. R. and C. J. Rothfels. 2023. Diversification models conflate likelihood and prior, and cannot be compared using conventional model-comparison tools. Systematic Biology 72:713–722.

Mulvey, L. P., M. R. May, J. M. Brown, S. Höhna, A. M. Wright, and R. C. Warnock. 2024. Assessing the adequacy of morphological models using posterior predictive simulations. Systematic Biology Page syae055.

Napoli, J. 2024. Resolving the "ontogeny problem" in vertebrate paleontology. bioRxiv .

Nylander, J. A., F. Ronquist, J. P. Huelsenbeck, and J. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53:47–67.

O'Reilly, J. E., M. N. Puttick, L. Parry, A. R. Tanner, J. E. Tarver, J. Fleming, D. Pisani, and P. C. Donoghue. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. Biology letters 12:20160081.

Parins-Fukuchi, C. 2018. Use of continuous traits can improve morphological phylogenetics. Systematic Biology 67:328–339.

Peters, S. E. and M. McClennen. 2016. The paleobiology database application programming interface. Paleobiology 42:1–7.

Pogue, M. G. and M. Mickevich. 1990. Character definitions and character state delineation: the bête noire of phylogenetic inference. Cladistics 6:319–361.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in bayesian phylogenetics using tracer 1.7. Systematic biology 67:901–904.

Scotland, R. W., R. G. Olmstead, and J. R. Bennett. 2003. Phylogeny reconstruction: the role of morphology. Systematic biology 52:539–548.

Sidor, C. A., N. J. Tabor, and R. M. Smith. 2021. A new late permian burnetiamorph from zambia confirms exceptional levels of endemism in burnetiamorpha (therapsida: Biarmosuchia) and an updated paleoenvironmental interpretation of the upper madumabisa mudstone formation. Frontiers in Ecology and Evolution 9:685244.

Smith, R., B. Rubidge, M. Day, and J. Botha. 2020a. Introduction to the tetrapod biozonation of the karoo supergroup. South African Journal of Geology 123:131–140.

Smith, R., B. Rubidge, and M. van der Walt. 2013. Therapsid biodiversity patterns and paleoenvironments of the karoo basin, south africa. Pages 31–62 *in* Forerunners of Mammals: Radiation, Histology, Biology (A. Chinsamy-Turan, ed.) Indiana University Press, Bloomington.

Smith, R. M. H., B. S. Rubidge, M. O. Day, and J. Botha. 2020b. Introduction to the tetrapod biozonation of the Karoo Supergroup. South African Journal of Geology 123:131–140.

Sulej, T. and G. Niedźwiedzki. 2019. An elephant-sized late triassic synapsid with erect limbs. Science 363:78–80.

Sullivan, C., R. R. Reisz, and R. M. Smith. 2003. The permian mammal-like herbivore diictodon, the oldest known example of sexually dimorphic armament. Proceedings of the Royal Society of London. Series B: Biological Sciences 270:173–178.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Some Mathematical Questions in Biology: DNA Sequence Analysis 17:57–86.

Thiele, K. 1993. The holy grail of the perfect character: the cladistic treatment of morphometric data. Cladistics 9:275–304.

Thorne, J., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution 15:1647–1657.

Tribble, C. M., W. A. Freyman, M. J. Landis, J. Y. Lim, J. Barido-Sottani, B. T. Kopperud, S. Höhna, and M. R. May. 2022. RevGadgets: An R package for visualizing Bayesian phylogenetic analyses from RevBayes. Methods in Ecology and Evolution 13:314–323.

Uhen, M. D., B. Allen, N. Behboudi, M. E. Clapham, E. Dunne, A. Hendy, P. A. Holroyd, M. Hopkins, P. Mannion, P. Novack-Gottshall, et al. 2023. Paleobiology database user guide version 1.0. PaleoBios 40.

Viglietti, P. A., R. M. Smith, K. D. Angielczyk, C. F. Kammerer, J. Fröbisch, and B. S. Rubidge. 2016. The daptocephalus assemblage zone (lopingian), south africa: a proposed biostratigraphy based on a new compilation of stratigraphic ranges. Journal of African Earth Sciences 113:153–164.

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

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

Wright, A. M. 2019. A systematist's guide to estimating Bayesian phylogenies from morphological data. Insect systematics and diversity 3:2.

Wright, A. M. and D. M. Hillis. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. PLoS One 9:e109210.

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

Wright, A. M. and B. M. Wynd. 2024. Modeling of rate heterogeneity in datasets compiled for use with parsimony. bioRxiv Pages 2024–06.

Wynd, B. M., B. R. Peecook, M. R. Whitney, and C. A. Sidor. 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. Journal of Vertebrate Paleontology 37:228–239.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. Journal of Molecular Evolution 39:306–314.

Zhang, R., A. J. Drummond, and F. K. Mendes. 2024. Fast bayesian inference of phylogenies from multiple continuous characters. Systematic Biology 73:102–124.

Zuckerkandl, E. and L. Pauling. 1962. Molecular disease, evolution, and genetic heterogeneity. Pages 189–225 *in* Horizons in Biochemistry (M. Kasha and B. Pullman, eds.) Academic Press, New York.