

# Bayesian estimation of fossil phylogenies and the evolution of early to middle Paleozoic crinoids (Echinodermata)

David F. Wright\*

School of Earth Sciences, The Ohio State University, Columbus, OH 43215, USA (wright.1433@osu.edu)

**Abstract.**—Knowledge of phylogenetic relationships among species is fundamental to understanding basic patterns in evolution and underpins nearly all research programs in biology and paleontology. However, most methods of phylogenetic inference typically used by paleontologists do not accommodate the idiosyncrasies of fossil data and therefore do not take full advantage of the information provided by the fossil record. The advent of Bayesian ‘tip-dating’ approaches to phylogeny estimation is especially promising for paleosystematists because time-stamped comparative data can be combined with probabilistic models tailored to accommodate the study of fossil taxa. Under a Bayesian framework, the recently developed fossilized birth–death (FBD) process provides a more realistic tree prior model for paleontological data that accounts for macroevolutionary dynamics, preservation, and sampling when inferring phylogenetic trees containing fossils. In addition, the FBD tree prior allows for the possibility of sampling ancestral morphotaxa. Although paleontologists are increasingly embracing probabilistic phylogenetic methods, these recent developments have not previously been applied to the deep-time invertebrate fossil record. Here, I examine phylogenetic relationships among Ordovician through Devonian crinoids using a Bayesian tip-dating approach. Results support several clades recognized in previous analyses sampling only Ordovician taxa, but also reveal instances where phylogenetic affinities are more complex and extensive revisions are necessary, particularly among the Cladida. The name Porocrinoidea is proposed for a well-supported clade of Ordovician ‘cyathocrines’ cladids and hybocrinids. The Eucladida is proposed as a clade name for the sister group of the Flexibilia herein comprised of cladids variously considered ‘cyathocrines,’ ‘dendrocrines,’ and/or ‘poteriocrines’ by other authors.

## Introduction

Modern macroevolutionary research resides at the nexus of paleontology and phylogenetic comparative biology. The fossil record provides a spectacular temporal window into the vicissitudes of life’s history, and paleontologists have long used its patterns to investigate large-scale trends in diversification dynamics and morphologic evolution over timescales inaccessible to experimental manipulation or field-based investigation (Simpson, 1944; Sepkoski, 1981; Hunt et al., 2008; Alroy, 2010). Similarly, biologists armed with molecular phylogenies of extant species and tree-based statistical techniques have increasingly become interested in addressing macroevolutionary questions traditionally studied by paleontologists (e.g., O’Meara et al., 2006; Bokma, 2008; Rabosky, 2009; Rabosky and McCune, 2009; Harmon et al., 2010; Pennell et al., 2014). Although differences between paleontologic and biologic perspectives remain, attempts to bridge disciplinary gaps between fields have wide-reaching implications for assembling a more synthetic macroevolutionary theory (Jablonski, 2008; Slater and Harmon, 2013; Hunt and Slater, 2016).

Instances of integration between fields, such as paleontology and molecular phylogenetics, often provide opportunities for reciprocal illumination. For example, fossils play a major role in dating

divergences among extant species. Without external information to constrain absolute ages, branch length estimation is confounded by the fact that both rates of molecular sequence evolution and elapsed time contribute to observed distances among species. Thus, the construction of a time-calibrated molecular phylogeny requires information on fossil morphologies and their temporal distributions to provide a numerical timescale for testing alternative models of macroevolutionary dynamics (Donoghue and Benton, 2007; dos Reis et al., 2016; Ksepka et al., 2015). Equally illuminating for paleontologists, many probabilistic methods originally developed by molecular phylogeneticists can be modified and applied to paleontologic data (Wagner, 2000a; Wagner and Marcot, 2010; Lee and Palci, 2015; but see Spencer and Wilberg, 2013). For example, Lewis (2001) developed a *k*-state Markov model for calculating likelihoods of discrete, morphologic characters based on a generalization of the Jukes-Cantor model of molecular sequence evolution. Although simplistic, Lewis’s (2001) ‘Mk’ model has recently been demonstrated in a Bayesian context to outperform other phylogenetic methods under a range of conditions present in real data sets, including missing character data, high rates of character evolution (and therefore homoplasy), and rate heterogeneity among characters (Wright and Hillis, 2014; O’Reilly et al., 2016). The recent resurgence of ‘total-evidence’ (Pyron, 2011; Ronquist et al., 2012) approaches in phylogenetics coincides with a renewed interest among biologists in phenotypic evolution and the utility of morphologic phylogenetics in an age of ‘post-molecular systematics’ (Lee and Palci, 2015; Pyron, 2015).

\* Present address: Department of Paleobiology, National Museum of Natural History, The Smithsonian Institution, P.O. Box 37012, MRC 121, Washington, DC 20013-7012, USA (wrightda@si.edu)

This revival of interest in morphologic phylogenetics is good news for paleontologists because nearly all phylogenies of fossil species are inferred using only morphologic character data. Indeed, there has been an increasing number of studies employing probabilistic approaches to estimate phylogenies with morphologic data, especially in paleontology (e.g., Wagner, 1998, 1999; Snively et al., 2004; Pollitt et al., 2005; Clarke and Middleton, 2008; Pyron, 2011; Ronquist et al., 2012; Slater, 2013, 2015; Wright and Stigall, 2013; Lee et al., 2014; Close et al., 2015; Bapst et al., 2016; Gorscak and O'Connor, 2016).

Particularly promising for systematic paleontology is the advent of tip-dating approaches for inferring phylogenies containing non-contemporaneous taxa (Pyron, 2011; Ronquist et al., 2012; Gavryushkina et al., 2014). Bayesian total-evidence tip-dating combines molecular sequences, morphologic character data, and temporal information on fossil distributions to simultaneously estimate the best tree topologies, branch lengths, and divergence times among extinct and extant lineages (Ronquist et al., 2012; Lee and Palci, 2015). Tip-dating approaches operate on the simple assumption that evolution can be modeled as a function of time, with either a strict or relaxed clock-like model of character change. Although most tip-dating studies combine fossil and living species (e.g., Pyron, 2011; Ronquist et al., 2012; Slater, 2013), tip-dating approaches equally apply to character matrices containing only morphologic data (Slater, 2015) and/or with only fossil taxa (Lee et al., 2014; Bapst et al., 2016; Gorscak and O'Connor, 2016). Moreover, mathematical models originally developed for studying the spread of viruses in epidemiology have found applications in fossil tip-dating (Stadler et al., 2012; Stadler and Yang, 2013; Gavryushkina et al., 2014). The ‘fossilized birth–death’ process (Stadler, 2010; Heath et al., 2014) has recently been applied within a Bayesian context as a more realistic tree prior distribution that accounts for macroevolutionary and sampling processes (Gavryushkina et al., 2014).

This paper presents the first application of Bayesian tip-dating methods to a fossil-only data set of invertebrate animals. Here, I examine phylogenetic relationships among early to middle Paleozoic crinoids (Echinodermata). Crinoids are particularly amenable for the purposes herein because: (1) they have a well-sampled fossil record (Foote and Raup, 1996); (2) their skeletal morphology is highly complex and character-rich (Ubaghs, 1978; Foote, 1994; Ausich et al., 2015); and (3) testing phylogenetic hypotheses among crinoid higher taxa requires sampling non-contemporaneous taxa over long timescales ( $>10^7$  years), making them an ideal system for implementing a tip-dating approach (Ronquist et al., 2012). Because the approach taken herein is novel to the invertebrate fossil record, I provide a brief discussion on Bayesian tip-dating and the fossilized birth–death process tree prior to familiarize the reader with these emerging methods. Although this makes the paper necessarily technical in places, it is hoped those sections will provide a useful resource for other systematic paleontologists interested in probabilistic approaches to fossil phylogenetics.

### Previous work on crinoid phylogeny

The Crinoidea form the sister group to all other extant echinoderm classes (Asteroidea, Echinoidea, Holothruoidea, and Ophiuroidea) and have an extensive geologic history spanning

the Lower Ordovician (ca. 480 Ma) to the present day. Ever since Bathér (1899) published his seminal work *A Phylogenetic Classification of the Pelmatozoa*, crinoid systematists have sought a robust evolutionary template for understanding the phylogenetic distribution of fossil and living species (Ausich and Kammer, 2001). Other than a few isolated studies conducted at low taxonomic levels (e.g., Kammer, 2001; Gahn and Kammer, 2002), most phylogenetic research using computational methods has focused on inferring relationships within two key time intervals: the Ordovician and the Recent (Ausich, 1998; Guensburg, 2012; Hemery et al., 2013; Rouse et al., 2013; Ausich et al., 2015; Summers et al., 2014; Cole, 2017). These intervals are significant because they bookend the evolutionary history of crinoids into their early diversification during the Ordovician Period and their present-day diversity in marine ecosystems. However, these intervals are separated by ~480 million years, and phylogenetic research linking post-Ordovician stem taxa with the crown Crinoidea remains a largely unexplored area of research (Simms, 1988; Simms and Sevastopulo, 1993; Webster and Jell, 1999).

Crinoids are traditionally divided into several higher taxa, including the Camerata, Disparida, Hybocrinida, Cladida, Flexibilia, and the Articulata (Moore and Teichert, 1978). Except for articulate crinoids, these groups first appear in Ordovician rocks. Despite more than a century of controversy, phylogenetic relationships among Ordovician taxa are reaching a consensus. For example, all recent analyses of Ordovician crinoids strongly support an early divergence between camerate and non-camerate crinoids (Guensburg, 2012; Ausich et al., 2015; Cole, 2017). Thus, the Camerata is the sister clade to all non-camerate crinoids. Similarly, both Guensburg (2012) and Ausich et al. (2015) recovered a monophyletic Hybocrinida as the sister clade to a subset of cladid taxa. Ordovician analyses also recovered a monophyletic Disparida as sister to the clade of cyathocrine cladids and hybocrinids (Guensburg, 2012; Ausich et al., 2015). However, relationships among taxa placed currently within the Cladida, and their relationships to other higher taxa, have been a long-standing problem in crinoid systematics (McIntosh, 1986, 2001; Sevastopulo and Lane, 1988; Kammer and Ausich, 1992, 1996; Simms and Sevastopulo, 1993).

In his review of echinoderm phylogeny and classification, Smith (1984) lamented the cladid portion of the crinoid tree was one of the “outstanding areas of ignorance in echinoderm phylogeny” (Smith, 1984, p. 456). Indeed, the Cladida (sensu Moore and Laudon, 1943) have long been considered a paraphyletic group because some nominal cladids are hypothesized to be more closely related to flexible and/or articulate crinoids than other cladids (Springer, 1920; Simms and Sevastopulo, 1993; Brower, 1995; Ausich, 1998; Wright, 2015). Unfortunately, recent phylogenetic analyses not only confirm that Ordovician cladids are a paraphyletic assemblage (Guensburg, 2012; Ausich et al., 2015), but also that the validity of the Cladida and their constituent higher taxa (i.e., Dendrocrinida and Cyathocrinida) cannot be fully remedied by simply adopting Simms and Sevastopulo’s (1993) recommendation to place the Flexibilia and Articulata within the Cladida. In addition, because the monophyletic status of a taxon requires a temporal reference frame (conventionally taken as the present day), it is unknown whether some recovered ‘clades’ in Ordovician analyses retain their monophyletic status when

younger taxa are considered. For example, Ordovician cyathocrine cladids are typically recovered as a clade (Ausich, 1998; Ausich et al., 2015) but are sometimes nested within a more inclusive clade of cyathocrines and hybocrinids when hybocrinids are sampled in the same analysis. Testing whether the other cyathocrine cladids belong to this clade remains an open question and requires sampling younger species. Similarly, Ordovician cladids placed within the Dendrocrinida (sensu Moore and Laudon, 1943) are paraphyletic, but there may nevertheless be latent phylogenetic structure present among subsets of post-Ordovician dendrocrines that could inform taxonomic revisions.

The analyses conducted herein build on and further test recently proposed phylogenetic hypotheses that sample only Ordovician crinoids (Guensburg, 2012; Ausich et al., 2015). This analysis includes a broad sample of early to middle Paleozoic (Ordovician–Devonian) non-camerate crinoids and primarily focuses on resolving relationships among the problematic Cladida.

### Bayesian phylogenetics and the fossilized birth–death process

Bayesian phylogenetic methods combine a likelihood model of evolution with a set of prior probabilities to generate a posterior probability distribution of phylogenetic trees and their associated parameters. The Bayesian framework used herein is adapted from Gavryushkina et al. (2015). This model uses time-stamped morphologic character data to simultaneously estimate a posterior distribution of phylogenetic trees, divergence times, and other macroevolutionary and sampling parameters.

Let  $\Psi$  be a phylogeny (i.e., tree topology with branch lengths in units of time),  $\delta$  be a vector of parameters describing morphologic evolution, and  $\pi$  be the tree prior (and its associated hyper parameters). Using Bayes theorem, the posterior probability distribution is

$$f[\Psi, \pi, \delta | X, d] = \frac{f[X | \Psi, \delta] f[d | \Psi] f[\Psi | \pi] f[\pi] f[\delta]}{f[X, d]}, \quad (1)$$

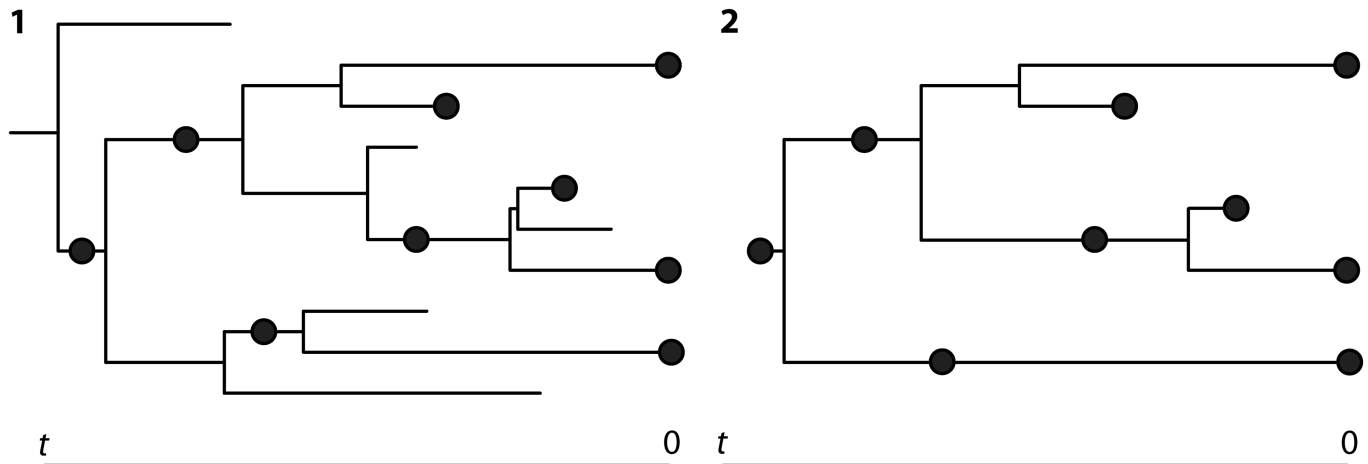
where  $X$  is a character by taxon matrix of morphologic character data and  $d$  is a vector of age ranges for each fossil taxon. The numerator on the right-hand side of the equation can be separated into the tree likelihood function,  $f[X | \Psi, \delta]$ , and the remaining terms, which comprises the prior. Equations for calculating  $f[X | \Psi, \delta]$  are well described in literature, and therefore a disquisition on tree likelihoods is not presented here. Interested readers are advised to see summaries in Swofford et al. (1996), Lewis (2001), Felsenstein (2004), and Yang (2014). The density  $f[\Psi | \pi]$  describes the tree prior (see below) and  $f[d | \Psi]$  is the density of obtaining fossil occurrence ranges given  $\Psi$  (this term is treated herein as a constant, see Gavryushkina et al., 2015). The denominator  $f[X, d]$  is a normalizing constant and is equal to the marginal probability of the data. Given the necessary inputs, the posterior distribution of trees is estimated using a numerical technique called Markov chain Monte Carlo (MCMC) that eliminates the need to calculate  $f[X, d]$  when estimating the posterior distribution of trees.

A great strength of the Bayesian paradigm is that sources of uncertainty can be explicitly incorporated into the evolutionary model via the use of prior distributions on pertinent parameters

(Heath and Moore, 2014). For example, numerous factors can influence and potentially distort the accuracy of reconstructed evolutionary trees—arguably the most important parameter in phylogenetic inference. Even in cases where these other factors are not of primary interest, acknowledging and estimating ‘nuisance parameters’ is nevertheless important because it reduces the chance that any particular incorrect assumption will lead to the recovery of specious tree topologies (Huelsenbeck et al., 2002; Wagner and Marcot, 2010; Gavryushkina et al., 2014). Potential biasing factors may include variation in rates of morphologic evolution, taxonomic diversification rates, ancestor–descendant relationships, and (incompletely) sampling taxa over time rather than from a single time slice (Smith, 1994; Wagner, 2000b, 2000c; Wagner and Marcot, 2010; Bapst, 2012). Variability in evolutionary rates can be modeled with prior distributions to describe rate variation among characters. Similarly, rate variation among lineages can be modeled using uncorrelated ‘relaxed clock’ models where branch-specific rates are independently drawn from the same underlying parametric distribution (Lepage et al., 2007; Heath and Moore, 2014).

Bayesian inference weights the likelihood of a tree by its prior probability. The fossilized birth–death (FBD) process (Stadler, 2010; Didier et al., 2012; Heath et al., 2014) is an extension of the constant rate birth–death models commonly used in paleontology (e.g., Raup et al., 1973; Raup, 1985) and considers fossil preservation in addition to diversification dynamics. In the following, I briefly describe the FBD process as a tree prior and argue it is well suited to accommodate these additional sources of concern.

*Tree prior.*—The fossilized birth–death (FBD) process is a stochastic branching model for describing macroevolutionary dynamics, fossil preservation, and sampling (Stadler, 2010; Heath et al., 2014). The FBD process begins at some time  $t_0 > 0$  in the past and ends when  $t = 0$ . As time moves forward (i.e., decreasing toward the ‘present’), each lineage may probabilistically undergo one of three process-based events, each according to a distinct constant rate Poisson process: branching (i.e., lineage splitting via speciation) with rate  $p$ , extinction with rate  $q$ , or fossil preservation and sampling with rate  $r$  (Stadler, 2010; Heath et al., 2014). Lineages alive at the end of the process are sampled with probability  $\varepsilon$  corresponding to the sampling fraction of ‘extant’ taxa. It is important to note that the start and end times for the FBD process are arbitrary, and time can therefore be shifted to accommodate temporal frameworks more commonly used in paleontology. For example, paleontological systematists working on entirely extinct groups (e.g., trilobites) or sampling taxa from a restricted temporal interval (e.g., Paleozoic crinoids) can shift time such that  $t = 0$  corresponds to the age of the youngest species sampled. The FBD process represents a major advance over other birth–death models in paleontology (e.g., Raup, 1985) because fossil preservation and sampling issues are modeled in addition to clade diversification. In the implementation of Gavryushkina et al. (2014), a lineage may be sampled more than once, thereby producing an internal node connected to only two (rather than three) branches. A two-degree internal node in a phylogenetic tree implies a hypothesized ancestor–descendent relationship, via direct or indirect ancestry (Fig. 1) (Foote, 1996; Gavryushkina et al., 2014, 2015).



**Figure 1.** Illustration of the fossilized birth–death process. (1) A full realization of the FBD process from time  $t$  in the past to the end of the process. Diversification produces a tree with branching and extinction events and random sampling of nodes. Sampling events are indicated by dots. (2) A ‘reconstructed’ phylogenetic tree produced by pruning all of the unsampled lineages in 1.1. Thus, only the observed portion of samples participating in the macroevolutionary process is depicted. Note that some sampled nodes represent ancestors.

The set of stochastic branching, extinction, and sampling events for a single FBD process gives rise to a ‘complete’ phylogeny with generating parameters  $\pi = (p, q, r, \varepsilon, t_0)$  (Fig. 1.1). The ‘sampled’ FBD phylogeny is obtained when all lineages with unsampled descendants produced by the process are pruned from the tree and therefore represents the reconstructed tree topology and divergence times implied by the sampled taxa (Fig. 1.2). Trees sampled from the FBD process are called sampled ancestor phylogenetic trees (even if no ancestors were sampled), and their nodes can be labeled to summarize their unique history of macroevolutionary and sampling events (Gavryushkina et al., 2014). Equations for calculating the probability density of  $f[\Psi|\pi]$  given the FBD parameters  $p, q, r$ , and  $\varepsilon$  can be derived by modifying birth–death sampling models used to study virus transmissions in epidemiology (Stadler, 2010; Stadler et al., 2012; Gavryushkina et al., 2014; Zhang et al., 2016), and their details are discussed in the Appendix.

### Taxon sampling, characters analyzed, and specimens examined

The character matrix analyzed herein was constructed as part of a larger project resolving phylogenetic relationships among Paleozoic crinoids. I updated, modified, and expanded the character list of Ausich et al. (2015) to include an ensemble of new characters to better capture variation among post-Ordovician taxa, particularly among ‘cladids’ (Ausich et al., 2015; Wright and Ausich, 2015). Because the taxonomic diversity of fossil crinoids is formidably high for comprehensive analysis, taxon sampling was restricted to Ordovician through Devonian species and multiple exemplars were sampled at pertinent taxonomic scales appropriate to the present analysis (Brusatte, 2010). The matrix was constructed in an attempt to maximize sampling across the broad spectrum of taxonomic, morphologic, and preservation gradients while keeping rigorous analysis tractable (Wagner, 2000c; Carlson and Fitzgerald, 2007; Heath et al., 2008).

The data set contains representative species from Ordovician, Silurian, and Devonian families of nominal ‘cladids’ (including cyathocrines and dendrocrines), disparids, hybocrinids, and flexibles (all taxa sensu Moore and Laudon, 1943). Species chosen as exemplars were typically the type species of a type genus that well characterizes the distribution of morphologic traits for each higher taxon, but sometimes geologically older species and/or more complete specimens were sampled instead (Table 1). Characters, plate homologies, and terminology are after Ubahgs (1978) and Ausich et al. (2015), with updates from Webster and Maples (2006) and Wright (2015). All but four traits were treated as unordered binary or multistate characters (Supplemental Data 1). These four characters were ordered based on known patterns of crinoid development, and arguments for ordering these traits are discussed by Wright (2015) and Webster and Maples (2006, 2008). Unknown and inapplicable character states were coded as missing. This new compilation of more than 3,000 specimen-based observations is the largest and most comprehensive morphologic data matrix ever constructed sampling Ordovician and post-Ordovician fossil crinoids (Supplemental Data 2).

In the final matrix, a total of 87 discrete morphologic characters comprising more than 300 character states were sampled across 42 species of non-camerate crinoids (Supplemental Data 2). Camerates were not included in the analysis because they diverged from non-camerate crinoids by at least the earliest Ordovician (Guensburg and Sprinkle, 2003; Guensburg, 2012; Ausich et al., 2015; Cole, 2017). Although tip-dating analyses do not per se require use of an outgroup (Ronquist et al., 2012), there are several reasons I used the Tremadocian species *Apektocrinus ubahgsi* Guensburg and Sprinkle, 2009 to assist rooting the tree. *Apektocrinus* was originally described as a tentative cladid that featured traits intermediate between protocrinoids and nominal cladids (Guensburg and Sprinkle, 2009). Protocrinoids were originally considered basal crinoids stemward of the divergence between camerates and non-camerates (Guensburg and Sprinkle, 2003). However, Guensburg (2012) subsequently placed protocrinoids within the Camerata, and later analyses by Ausich et al. (2015) recovered protocrinoids to be closer to non-camerates than to camerates. Regardless of the labile phylogenetic position of

**Table 1.** Species sampled for phylogenetic analysis.

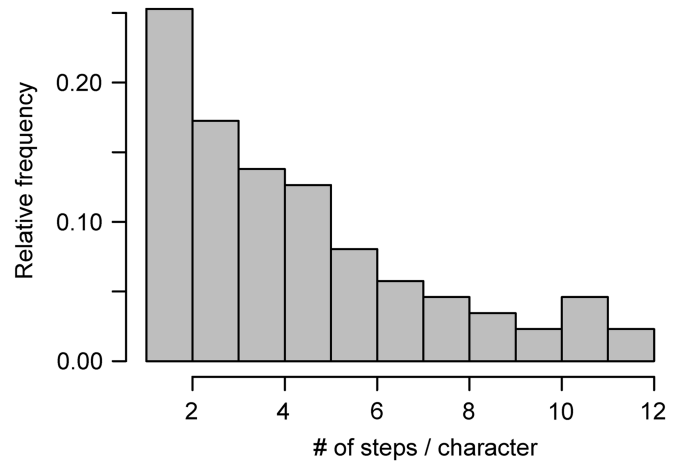
Species	Author
<i>Aethocrinus moorei</i>	Ubaghs, 1969
<i>Alphacrinus mansfieldi</i>	Guensburg, 2010
<i>Amabilicrinus iranensis</i>	Webster, Maples, Mawson, and Dastanpour, 2003
<i>Apektocrinus ubaghsi</i>	Guensburg and Sprinkle, 2009
<i>Botryocrinus ramosissimus</i>	Angelin, 1878
<i>Carabocrinus radiatus</i>	Billings, 1857
<i>Codiocrinus granulatus</i>	Schultze, 1867
<i>Colpodeocrinus quadrifidus</i>	Sprinkle and Kolata, 1982
<i>Corematocrinus plumosus</i>	Goldring, 1923
<i>Crotalocrinites verucosus</i>	(Schlotheim, 1820)
<i>Cupulocrinus humilis</i>	(Billings, 1857)
<i>Dendrocrinus longidactylus</i>	Hall, 1852
<i>Euspirocrinus spiralis</i>	Angelin, 1878
<i>Eustenocrinus springeri</i>	Ulrich, 1925
<i>Gasterocoma antiqua</i>	Goldfuss, 1839
<i>Glossocrinus naplesensis</i>	Goldring, 1923
<i>Heviacrinus melendezi</i>	Gil Cid, Alonso, and Pobes, 1996
<i>Homalocrinus nanus</i>	(Salter, 1873)
<i>Hybocrinus conicus</i>	Billings, 1857
<i>Hybocystites problematicus</i>	Wetherby, 1880
<i>Ibexocrinus leptus</i>	Lane, 1970
<i>Icthyocrinus laevis</i>	Conrad, 1842
<i>Iocrinus subcrassus</i>	(Meek and Worthen, 1865)
<i>Lecanocrinus macropetalus</i>	Hall, 1852
<i>Lecythocrinus eifelianus</i>	Müller, 1859
<i>Manicrinus hybocriniformis</i>	Frest and Strimple, 1978
<i>Mastigocrinus arboreus</i>	(Salter, 1873)
<i>Merocrinus typus</i>	Walcott, 1884
<i>Metabolocrinus rossicus</i>	Jaekel, 1902
<i>Ottawacrinus typus</i>	Billings, 1887
<i>Petalocrinus mirabilis</i>	Weller and Davidson, 1896
<i>Plicodendrocrinus casei</i>	Meek, 1871
<i>Porocrinus conicus</i>	Billings, 1857
<i>Proctothylacocrinus longus</i>	Kier, 1952
<i>Protaxocrinus ovalis</i>	(Angelin, 1878)
<i>Rhenocrinus ramoississimus</i>	Jaekel, 1906
<i>Rutkowskicrinus patriciae</i>	McIntosh, 2001
<i>Sagenocrinites expansus</i>	(Phillips, 1839)
<i>Sphaerocrinus geometricus</i>	(Goldfuss, 1831)
<i>Streblocrinus brachiatus</i>	Koenig and Meyer, 1965
<i>Thalamocrinus ovatus</i>	Miller and Gurley, 1895
<i>Thenarocrinus callipygus</i>	Bather, 1890

protocrinids, both Guensburg (2012) and Ausich et al. (2015) recovered tree topologies with *Apektocrinus* as the sister taxon to the clade comprised of non-camerate crinoids. Thus, the early stratigraphic position, mosaic distribution of plesiomorphic and apomorphic traits, and strong support from previous phylogenetic analyses all indicate *Apektocrinus* occupies a position near the base of the non-camerate tree (Guensburg and Sprinkle, 2009; Guensburg, 2012; Ausich et al., 2015).

Matrix construction required extensive first-hand examination of well-preserved specimens housed in museum collections and of the published taxonomic literature. When possible, I coded characters from direct observations of type-series specimens for each species. Although emphasis was placed on observing characters from type specimens, non-type specimens were also examined to ensure the character distributions for each species were coded as completely as possible. Specimens were examined from collections within the United States National Museum of Natural History, the Field Museum of Natural History, the Lapworth Museum of Geology, and the Natural History Museum (London).

### Phylogenetic analyses

Bayesian phylogenetic analyses were conducted using Markov chain Monte Carlo (MCMC) sampling in the MPI-version of



**Figure 2.** Parsimony-based estimate of rate variation among characters. This distribution suggests that many characters evolve slowly, whereas a small number of characters evolve at much higher rates.

MrBayes 3.2.5 (Ronquist et al., 2012), which implements MCMC proposals for FBD trees (Zhang et al., 2016). To account for differences among alternative model configurations, multiple phylogenetic analyses were conducted and Bayes factors (BFs) were calculated to statistically compare models. Bayes factors are used in Bayesian model selection to determine which parameter configurations provide the best fit to the data and are equal to twice the difference in marginal log-likelihoods between models (Kass and Raftery, 1995). Following phylogenetic analyses, I then estimated the marginal log-likelihood of each model using the stepping-stone sampling method (Xie et al., 2010) with 50 steps and powers of  $\beta$  corresponding to quantiles of a Beta(0.5, 1.0) distribution. Parsimony-based calculations were performed using PAUP\* 4.0a147 (Swofford, 2002). All additional analyses were conducted using custom scripts written in the R statistical computing environment making use of functions from the packages APE (Paradis et al., 2004), and STRAP (Bell and Lloyd, 2015). Details regarding choices of prior distributions, constraints, and MCMC convergence are discussed in the following.

Morphologic character evolution was modeled using the Mk model with equal transition frequencies among character states and a correction for ascertainment bias in character acquisition (Lewis, 2001). The distribution of rates among characters can assume either a uniform ‘equal rates’ model or explicitly account for rate heterogeneity using a skewed parametric distribution. A preliminary parsimony-based estimation of rate variation in the crinoid character matrix depicts a highly skewed distribution (Fig. 2), strongly suggesting it is unwise to assume a model of equal rates of change among characters. This is particularly striking given that parsimony-based rate distributions tend to underestimate morphologic changes and are therefore slightly biased toward equal rates (Harrison and Larsson, 2015). To further test this hypothesis, I conducted separate analyses assuming equal, lognormal, and gamma distributed rates of character change. Following Harrison and Larsson’s (2015) recommendation, analyses with gamma or lognormal variation used eight instead of four discrete rate categories (commonly applied to molecular data).

Fossil tip-dating requires a model characterizing the distribution of evolutionary rates throughout the tree. The case of a strict-morphological clock assumes rates are constant among lineages. However, the assumption of the strict-clock can be ‘relaxed’ by allowing rates to vary among branches throughout the tree (Lepage et al., 2007; Heath and Moore, 2014). To test whether evolutionary rates vary among lineages, strict- and relaxed-clock analyses were conducted. The independent gamma rates (IGR) relaxed-clock model was applied to account for variation in rates among branches. The IGR model facilitates episodic ‘white noise’ variation in rates across the tree and is appropriate because large-scale morphologic evolution is a function of both waiting times and stochastic selective forces (Wagner, 2012; Heath and Moore, 2014). A lognormal distribution was placed on the base rate of the clock using methods outlined by Ronquist et al. (2012).

A key assumption of tip-dating is that evolutionary change is a function of time. In other words, geologically younger species are expected to have undergone a greater amount of within-lineage evolution (i.e., anagenesis) than older species because more time has elapsed for changes to occur (Smith et al., 1992; Wagner, 2000a). To test whether this assumption holds (and therefore whether the tip-dating method is valid for these data), the parsimony-based root-to-tip path length of each species from a non-clock analysis was regressed against median age dates from the IGR analysis using both phylogenetically corrected and uncorrected methods (Lee et al., 2014).

The FBD process was used as a tree prior (i.e., ‘samples-trat = random’ in MrBayes 3.2.5). The implementation of the FBD in MrBayes reparameterizes the FBD process in terms of net diversification ( $= p - q$ ), turnover ( $= q/p$ ), and sampling probability ( $= r/(q + r)$ ). I placed an Exp(1) prior on net diversification, a Beta(1,1) uniform prior on turnover, and a Beta(2,2) prior on the sampling probability. To account for uncertainty in divergence time estimation, age ranges for fossil species were given broad uniform distributions typically corresponding to the stratigraphic range of their higher taxon and were taken from an updated version of Webster’s (2003) index of Paleozoic crinoids (Supplemental Data 2). Because the age of the most recent common ancestor of all species in the analysis is well constrained by fossil evidence to be near the base of the Ordovician, the tree age prior was fixed to correspond to the earliest Tremadocian.

Tip-dating is a computationally demanding phylogenetic method that requires a time-consuming exploration of parameter space. To assist the analysis, several topological constraints were applied to reduce MCMC exploration of very unlikely trees and to test more specific phylogenetic hypotheses (see Guillerme and Cooper, 2016). For example, the monophyly of disparids and flexibles are well supported by other studies (Brower, 1995; Ausich, 1998; Guensburg, 2012; Ausich et al., 2015) and preliminary analyses. Thus, their status as clades is not in question. However, the branching position of these clades within the larger crinoid tree remains an open question, and their phylogenetic placement is evaluated herein. A partial constraint was placed on *Eustenocrinus*, *Iocrinus*, *Ibexocrinus*, and *Heviacrinus* that allowed for either *Merocrinus* and/or *Alphacrinus* to be included within the disparid clade if

the data support that hypothesis. This was done because whether *Merocrinus* is closer to cladids or disparids requires additional testing (cf. Ausich, 1998; Guensburg, 2012). *Alphacrinus* is a stratigraphically old taxon with a combination of unique and disparid-like traits that may or may not be stemward to ingroup Disparida (Guensburg, 2010). A hard constraint was placed on a flexible clade comprising *Homalocrinus*, *Ichthyocrinus*, *Lecanocrinus*, *Protaxocrinus*, and *Sagenocrinites*.

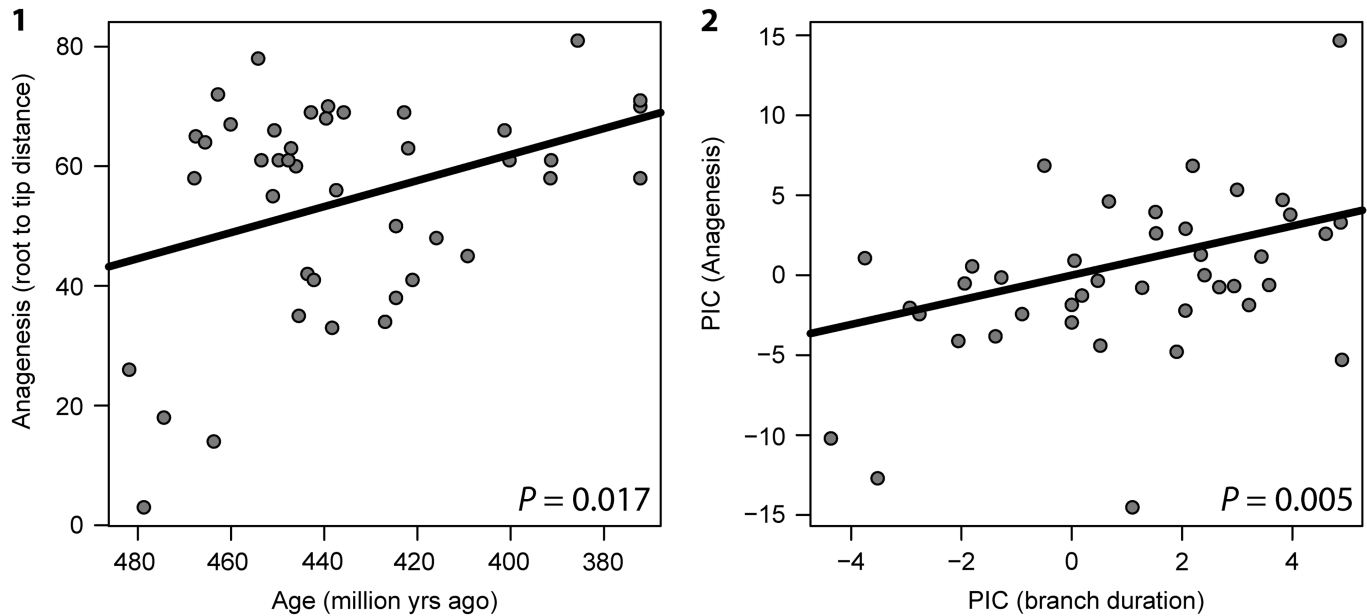
Markov chain Monte Carlo analyses consisted of two independent runs of four chains sampling every 4,000 generations for 40 million generations per run with a burn-in percentage of 35%. Convergence was assessed using multiple criteria: average standard deviation of split frequencies among chains were below 0.01 ( $<0.05$  for some IGR analyses) (Gelman and Rubin, 1992), potential scale reduction factors of  $\sim 1.0$  (Lakner et al., 2008), effective sample sizes greater than 300 (with many  $>1,000$ ), and visual inspection of log-likelihood plots among runs using Tracer v.1.6 (Drummond and Rambaut, 2007). Finally, the analysis with the best fit parameter settings was repeated three times to ensure estimates of optimal tree topologies were robust across runs. Together, these diagnostics indicate convergence among tree topologies and parameter estimates.

To summarize the posterior distribution of tree topologies, I generated a maximum clade credibility (MCC) tree using TreeAnnotator (Rambaut and Drummond, 2015). Although there is no single agreed-upon method for summarizing Bayesian posterior distributions of phylogenetic trees (Heled and Bouckaert, 2013), posterior probability can be viewed as an optimality criterion in phylogenetic inference (Rannala and Yang, 1996; Huelsenbeck et al., 2002; Wheeler and Pickett, 2007; Wheeler, 2012; Rambaut, 2014). The MCC tree is the tree in the posterior distribution with the maximum product of clade posterior probabilities and represents a Bayesian point estimate of phylogeny (Rambaut, 2014).

I also ran a series of sensitivity analyses ( $n > 5$ ) to explore the effects of choosing different priors, including the prior placed: (1) on the variance of the gamma distribution in the IGR model, and (2) on the FBD turnover ( $= q/p$ ) parameter. In all cases, statistically indistinguishable median estimates were obtained for node ages, branch lengths, and FBD parameters. In addition, I calculated the pairwise Robinson-Foulds (RF) distance (Robinson and Foulds, 1981) within and among tree distributions from separate analyses and ordinated the resulting RF matrix using principal coordinate analysis. A visual inspection of RF distances in principal coordinate space reveals substantial overlap between distributions, with no obvious gradient or isolated ‘islands’ (analogous to Maddison, 1991) of trees. Thus, the analysis presented herein is considered robust across a range of possible prior configurations.

## Results

The relationship between within-lineage morphologic evolution and IGR age estimates indicates early to middle Paleozoic crinoids conform well to the assumptions of tip-dating. Parsimony-based branch lengths from a non-clock analysis reveals geologically younger taxa have higher amounts of anagenesis compared to older taxa ( $p = 0.017$ ) (Fig. 3). This relationship holds even when



**Figure 3.** Testing statistical associations between the amount of morphologic evolution inferred from an undated analysis and divergence times estimated under a relaxed morphologic-clock model: (1) node age in millions of years and anagenesis (measured as the total root-to-tip distance in parsimony steps from an undated analysis); (2) phylogenetic independent contrasts between branch durations (from time-scaled analysis) and anagenesis (from an undated analysis). Both regressions are statistically significant.

accounting for phylogenetic non-independence among comparisons, as phylogenetic independent contrasts (Felsenstein, 1985) also reveal a strong statistical association between branch durations (in the best-fit IGR analysis) and morphologic divergence (measured in parsimony steps from an undated analysis) ( $p = 0.005$ ) (see Lee et al., 2014).

Bayes factors provide evidence for heterotachy in crinoid evolution throughout this interval (Supplemental Data 3). The equal rates model of character evolution was strongly rejected in favor of models incorporating rate variation (BF = 166.42 for lognormal, BF = 162.04 for gamma), with the lognormal slightly outperforming a gamma distribution (BF = 4.38). Similarly, the strict-morphological clock was strongly rejected in favor of the IGR relaxed-clock model accounting rate variation among lineages (BF = 54.56).

The MCC tree from the best fit model is presented in Figure 4. Although posterior probabilities for some clades are low, they are comparable to other tip-dating studies using morphologic characters (Lee et al., 2014; Gavryushkina et al., 2015; Gorscak and O'Connor, 2016). Because Bayesian analyses account for uncertainty in the phylogenetic placement of taxa, it is important to stress that relationships depicted in Figure 4 are not the only ones supported in the posterior distribution. However, MCC tree topologies were broadly consistent across all analyses, indicating topological results are robust across different model configurations. Keeping uncertainty in mind, there are many salient features of the MCC tree that support previous phylogenetic hypotheses and throw light on several taxonomic questions.

The base of the MCC tree is characterized by a basal divergence between disparids (and disparid-like taxa) and all other non-camerate crinoids. Neither *Merocrinus* nor *Alphacrinus* were placed within the clade comprising *Iocrinus*, *Ibexocrinus*, *Eustenocrinus*, and *Heviacrinus*. However, *Merocrinus* is placed as the sister taxon to *Metabolocrinus*, which together form a sister

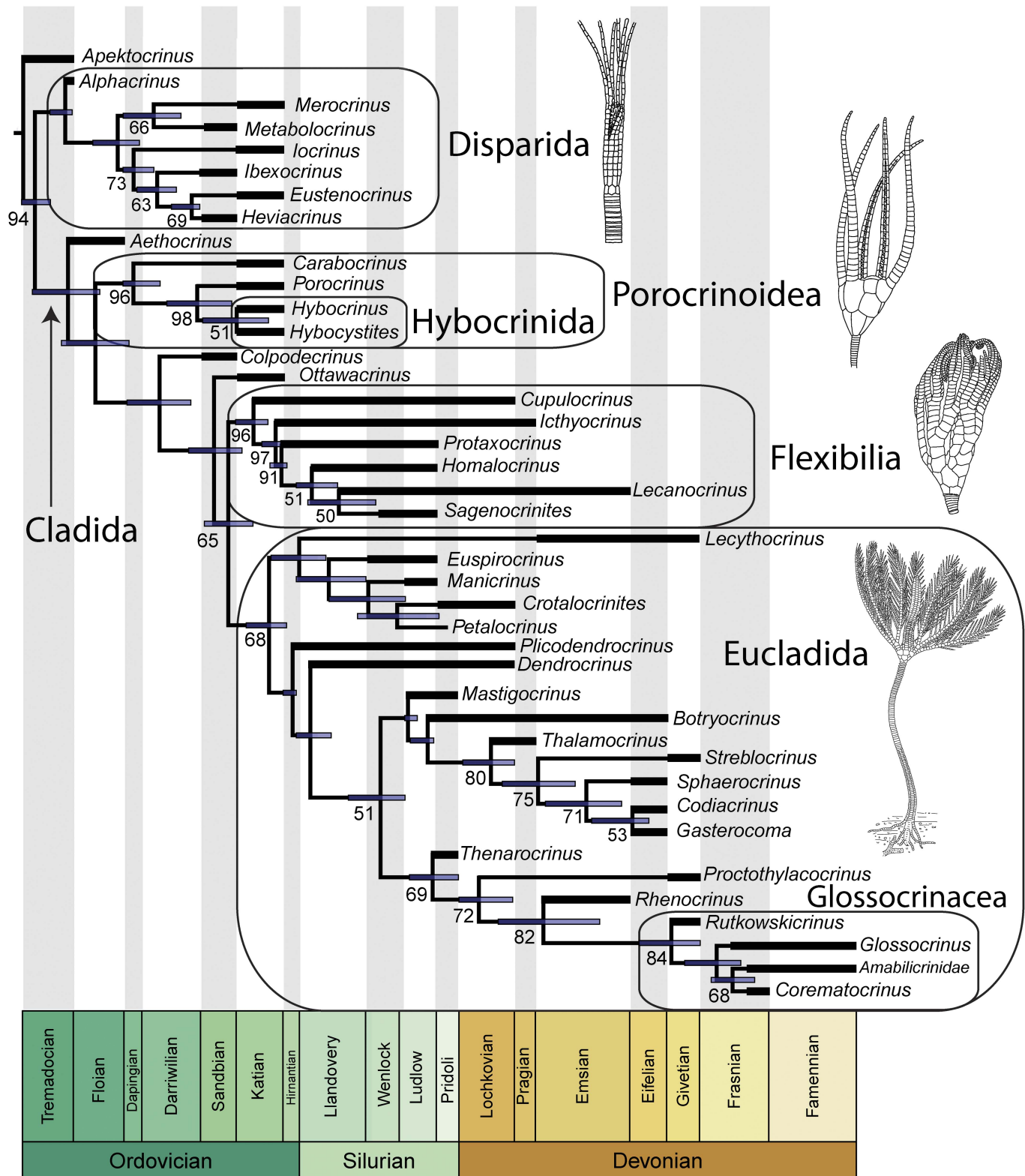
clade to the above-mentioned disparid clade. It is interesting to note that *Alphacrinus* is placed as the sister taxon to the clade comprised of *Merocrinus*, *Metabolocrinus*, and disparids, which supports Guensburg's (2010) original hypothesis of *Alphacrinus* being a basal disparid-like taxon and contrasts with Guensburg's (2012) subsequent analysis recovering *Alphacrinus* as nested within the disparid clade.

Similar to other studies (Ausich, 1998; Guensburg, 2012; Ausich et al., 2015), a clade comprised of cladids and hybocrinids was recovered. This clade is strongly supported with posterior probability 0.96. The hybocrinids *Hybocrinus* and *Hybocystites* are sister taxa and occupy a nested position within a clade of dicyclic 'cyathocrine' cladids (i.e., *Carabocrinus* and *Porocrinus*), suggesting these taxa are pseudomonocyclic (Sprinkle, 1982; Guensburg, 2012; Ausich et al., 2015).

A clade comprised of *Cupulocrinus* and flexible crinoids was recovered as sister to a clade comprised of only cladids, with the inclusive clade supported with posterior probability 0.65. Although *Cupulocrinus* is a nominal cladid (sensu Moore and Laudon, 1943), it is placed closer to flexibles than other cladids in 96% of trees in the posterior distribution.

The large clade comprising the most recent common ancestor of *Plicodendrocrinus* and *Corematocrinus* and its descendants contains a scattering of taxa traditionally placed within the cladid orders Cyathocrinida and Dendrocrinida (sensu Moore and Laudon, 1943). Thus, the traditionally recognized 'cyathocrine' and 'dendrocrine' cladids represent evolutionary grades of body plan organization rather than clades. For example, a clade of cyathocrine-grade crinoids containing the most recent common of *Thalamocrinus* and *Gasterocoma* is supported with posterior probability 0.80, with subclades supported by posterior probabilities between 0.53 and 0.75. Similarly, a different clade containing the most recent common ancestor of *Lecythocrinus* and *Petalocrinus* also





**Figure 4.** Maximum Clade Credibility tree of early to middle Paleozoic crinoids. Posterior probabilities (>.50) are located next to nodes and expressed in percent; blue node bars represent the 95% highest posterior density age estimates; thick black bars represent genus-level stratigraphic ranges. Note the inclusion of stratigraphic ranges gives the appearance of sister taxon relationships where zero-length branches were sampled (e.g., *Cupulocrinus*). The Cladida (sensu Simms and Sevastopulo, 1993) are sister to the Disparida. Crinoid taxa depicted represent the major clades discussed in the text: (from top to bottom) the disparid *Eustenocrinus springeri* Ulrich, 1925, redrawn from Ubahgs (1978); representative porocrinoid and hybocrinid (see text), *Hybocrinus conicus* Billings, 1857, redrawn from Sprinkle and Moore (1978); the flexible *Protaxocrinus laevis* (Billings, 1857), redrawn from Springer (1911); representative euclidid *Dictenocrinus*, redrawn from Bather (1900).



features ‘cyathocrine’ morphologies. These results support (perhaps unfortunately) long-held suspicions of taxonomic anarchy among cladids recognized by previous authors (McIntosh, 1986, 2001; Kammer and Ausich, 1992, 1996; Simms and Sevastopulo, 1993; Webster and Maples, 2006).

The clade of dendrocrine-grade crinoids containing the most recent common ancestor of *Thenarocrinus* and *Corematocrinus*, and its constituent subclades, are supported with posterior probabilities between 0.68 and 0.84. This clade contains a subclade comprising members of the Rutkowskiicrinidae, Glossocrinidae, Corematocrinidae, and Amabilicrinidae. This clade is supported with posterior probability 0.84 and is equivalent to the superfamily Glossocrinacea originally recognized by Webster et al. (2003).

## Discussion

It is generally appreciated that quantitative phylogenetic methods do not typically take full advantage of the complete spectrum of information supplied by the fossil record (Wagner and Marcot, 2010). However, recently developed probabilistic macroevolutionary models and powerful computational tools have provided major advancements for estimating phylogenies containing fossil taxa and accommodating paleontologic idiosyncrasies, such as sampling taxa (incompletely) over time (Stadler, 2010; Ronquist et al., 2012; Gavryushkina et al., 2014; Lee and Palci, 2015). This paper builds on these advances by implementing a Bayesian framework to estimate time-scaled phylogenetic hypotheses of early to middle Paleozoic fossil crinoids. The resulting phylogeny indicates extensive taxonomic revisions are necessary, especially among the ‘cladid’ crinoids, and points to several areas where further analysis at lower taxonomic levels are needed. In addition to evolutionary implications for crinoids, the results raise several key issues regarding probabilistic approaches to phylogenetic inference in the fossil record and suggest possible directions for future research.

*Implications for crinoid evolution and systematics.*—The phylogenetic analysis presented herein offers several insights into early to middle Paleozoic crinoid evolution and provides a basis for requisite taxonomic revisions. Although I provide suggestions for revisions below, attempting to more fully resolve outstanding problems in crinoid systematics and classification is beyond the scope of this paper. Instead, that topic is addressed in a companion paper (see Wright et al., 2017).

A basal divergence between disparids and most other non-camerate crinoids has been recovered in a number of recent phylogenetic studies (cf. Guensburg, 2012; Ausich et al., 2015), including the analysis herein. Although several nominal ‘cladids’ are stemward of this split (see Ausich et al., 2015, fig. 5), the overwhelming majority of nominal cladids (sensu Moore and Laudon, 1943) are not. Thus, disparids are nested within a clade comprising the common ancestor of all nominal ‘cladids’ (sensu Moore and Laudon, 1943) and all of its descendants. Following Simms and Sevastopulo (1993), the Flexibilia and Articulata are placed within the Cladida, but no previous phylogenetic hypothesis has considered the Disparida a subclade within the Cladida. In an effort to retain as much of

the original intent and traditional use of taxonomic names as possible, a redefinition of the Cladida is necessary to prevent the Disparida from being considered a subclade of cladids, particularly since the Cladida is already in need of extensive revision for other reasons. A simple solution to remedy the problem could be obtained using phylogeny-based clade definitions that recognize the Disparida and Cladida (sensu Simms and Sevastopulo, 1993) as sister clades. This would require orphaning only a small number of so-called cladids (sensu Moore and Laudon, 1943) as stem taxa to the Disparida + Cladida clade and retain the majority of cladids (sensu Moore and Laudon, 1943) within the more inclusive Cladida (sensu Simms and Sevastopulo, 1993). The Disparida is considered herein to include the most recent common ancestor of *Alphacrinus* and *Eustenocrinus*. Thus, *Merocrinus* and *Metabolocrinus* (typically considered cladids; Ausich, 1998; but see Guensburg, 2012) are tentatively placed within the disparid clade (Fig. 4), but further work and character-based analyses are needed to confirm the precise phylogenetic position of these problematic taxa. A more comprehensive discussion with rigorous phylogenetic definitions and a revised classification for cladid and disparid clades is provided in Wright et al. (2017).

Previous analyses of Ordovician taxa have recovered monophyletic groups, such as a clade comprised of cyathocrine cladids and hybocrinids (Guensburg, 2012; Ausich et al., 2015). This offered some hope that potentially the Cyathocrinida might be monophyletic (Ausich, 1998; Guensburg, 2012; Ausich et al., 2015). However, the present analysis rejects the monophyly of the Cyathocrinida because some nominal cyathocrines are more closely related to nominal dendrocrines than to other cyathocrines. If these results are taken seriously, then an extensive revision of higher taxa within the Cladida is needed. To further test this issue, I conducted an additional analysis placing a hard topological constraint on all cyathocrine cladids (see Bergsten et al., 2013). Comparing this analysis to the best fit model described above, Bayes factors strongly reject a model where cyathocrines are forced to be monophyletic (BF = 9.64). Thus, caution should be exercised when extrapolating results from an analysis considering one timeslice to subsequent time intervals. Nevertheless, there is strong support for a clade of cyathocrine-grade cladids and hybocrinids (Ausich, 1998; Guensburg, 2012; Ausich et al., 2015). This clade is characterized by a number of morphologic features convergent with blastozoan echinoderms, such as thecal respiratory structures, calyx and/or arm plate reduction, and recumbent ambulacra. Given the strong statistical support and morphologic distinctness of this clade, I propose the name Porocrinioidea to represent this idiosyncratic group of crinoids. The Hybocrinida is considered herein a subclade within the Porocrinioidea (Fig. 4).

The position of *Cupulocrinus* at the base of the flexible clade has strong statistical support and corroborates earlier studies linking *Cupulocrinus* with flexible crinoids (Springer, 1911, 1920; Brower, 1995; Ausich, 1998) (Fig. 4). For example, Springer (1920) considered *Cupulocrinus* to have traits intermediate between cladids and flexibles and hypothesized a species of *Cupulocrinus* was ancestral to the Flexibilia. Because phylogenetic relationships were estimated using methods that include the possibility of potentially sampling ancestral

morphotaxa, Springer's (1920) hypothesis can be quantitatively addressed. The posterior probability of a taxon being a sampled ancestor can be estimated as the frequency in which it was recovered to have a zero-length branch in the posterior distribution of trees (Matzke, 2015). Examining the posterior distribution of trees from the best-fit model, the probability of *Cupulocrinus humilis* (Billings, 1857) being an ancestral morphotaxon is 0.99. Given that a clade is defined to be an ancestor and all of its descendants, *Cupulocrinus* is removed from the Cladida and placed within the Flexibilia. Additional analyses with more comprehensive sampling of *Cupulocrinus* and flexible species (including a broader sample of taxon-specific characters) are needed to further test this hypothesis at finer taxonomic scales.

The sister clade to the Flexibilia contains the majority of all nominal taxa currently placed within the Cladida (sensu Moore and Laudon, 1943). This clade originated prior to the close of the Ordovician and contains most taxa traditionally placed within the orders Dendrocrinida and Cyathocrinida. Note that all species in this clade share a more recent common ancestor with an extant crinoid than with flexible crinoids (Simms and Sevastopulo, 1993). Thus, I propose the name Eucladida to distinguish this important group of crinoids from its sister clade.

The recovery of the Glossocrinacea as a clade provides quantitative support for evolutionary inferences discussed by Webster et al. (2003). These 'transitional dendrocrinids' (McIntosh, 2001) are among the first cladids to evolve pinnules and are traditionally placed within the order Poteriocrinida. Most crinoid workers since publication of the *Treatise of Invertebrate Paleontology* (Moore and Teichert, 1978) have hesitated to recognize the Poteriocrinida because they are widely considered to be polyphyletic (Kammer and Ausich, 1992; McIntosh, 2001). Regardless of their status as a clade or a grade, the crinoids considered poteriocrines in the *Treatise* are the most dominant and ecologically abundant group of crinoids throughout the middle to late Paleozoic (Ausich et al., 1994). Note that the ancestor of extant articulate crinoids is widely considered to be placed among a paraphyletic group of 'poteriocrine' crinoids (Simms and Sevastopulo, 1993; Webster and Jell, 1999; Rouse et al., 2013). The recovery of a clade of poteriocrines herein suggests there may be some phylogenetic structure present among Paleozoic poteriocrine taxa. However, future analyses sampling younger taxa and a broader sample of *Treatise* (Moore and Teichert, 1978) poteriocrines are needed to test whether this is the case.

*Probabilistic approaches to fossil phylogenies.*—Tree-based comparative methods are becoming commonplace in paleontology for testing macroevolutionary patterns and processes within a fully phylogenetic context. To date, most of these studies apply an a posteriori timescaling algorithm to an unscaled cladogram (e.g., Brusatte et al., 2008; Hunt and Carrano, 2010; Lloyd et al., 2012; Hopkins and Smith, 2015). Although useful for removing the zero-length branches that arise from polytomies in cladistic hypotheses, many of these a posteriori timescaling methods are problematic because they make ad hoc and unrealistic assumptions regarding node ages and/or ancestor–descendant relationships (Bapst and Hopkins, in press). The *cal3* method developed by Bapst (2013) is a

promising a posteriori approach that overcomes many of these problems via a model of branching, extinction, and sampling similar to the FBD process. However, this technique requires a priori estimates of these parameters and can only be applied to unscaled cladograms. The Bayesian tip-dating approach advocated herein simultaneously estimates tree topologies and divergence dates using time-stamped comparative data. Thus, a sample of trees from the posterior distribution of a tip-dated analysis provides a more natural framework for testing macroevolutionary patterns using the fossil record while accounting for uncertainty in tree topology and node ages (Close et al., 2015; Gorscak and O'Connor, 2016).

Evaluating the efficacy of competing phylogenetic methods is a contentious (and sometimes acrimonious) debate, yet inferences using simple probabilistic methods perform well when inferring trees from paleontologic data and can explicitly consider different evolutionary and sampling parameters potentially influencing recovered topologies (Wagner, 1998; Wagner and Marcot, 2010; Wright and Hillis, 2014). Thus, it seems that in the future, more phylogenetic analyses will take advantage of fully probabilistic frameworks such as the one presented herein. However, researchers conducting tree-based comparative analyses on older (or otherwise unscaled) cladograms require an a posteriori time-scaling approach, which might take the form of applying FBD-like divergence dating methods to a fixed cladistic topology (Bapst and Hopkins, in press). Regardless of whether Bayesian tip-dating or a model-based a posteriori method like *cal3* becomes the dominant approach to timescaling trees in the future, it is apparent that both of these approaches recover more accurate estimates and are strongly recommended over ad hoc methods when conducting downstream macroevolutionary analyses.

Evolutionary patterns among early to middle Paleozoic crinoids strongly favor models incorporating rate heterogeneity among characters and among lineages. This not only supports previous investigations demonstrating differential disparity patterns among crinoid clades (Foote, 1994; Deline and Ausich, 2011), but may also be a more general feature of morphologic evolution. Probabilistic models of morphologic evolution commonly assume either uniform or gamma distributed rates among characters. Models of rate variation predict some characters evolve at higher rates than others (and therefore anticipate a degree of homoplasy in the data). Variable rate distributions are potentially more realistic than an equal rates model because morphologic characters commonly experience different selective pressures and/or developmental constraints. Moreover, accounting for rate variation has a practical value because it may help resolve branches at different levels in a phylogenetic tree (Wright and Hillis, 2014).

Until recently, only uniform and gamma distributions were available to model rate variation among characters in common software packages for Bayesian inference (Huelsenbeck et al., 2015). However, Wagner (2012) found that fossil data sets commonly favor lognormal rather than gamma distributed rates, particularly for echinoderm and mollusk character matrices. It is interesting to note that gamma and lognormal rate distributions arise from different underlying processes of character evolution. Gamma rate distributions assume rates are Poisson processes, whereas lognormal rate distributions suggest morphologic

evolution results from multiple probabilistic processes and/or hierarchical integration among characters (Wagner, 2012). Thus, the relative fit of data to these two distributions gives some insight into the dynamics of character change. The analysis herein provides positive evidence supporting lognormal over gamma-distributed rates, but only modestly ( $BF = 4.38$ ). Although the choice of rate distribution had no obvious effect on recovered topologies herein, other workers should nevertheless test alternative distributions and choose the best-fit model for their data (Harrison and Larsson, 2015).

The FBD process provides paleontologists with a far more realistic tree prior model than others previously available. For example, the FBD tree prior has recently been demonstrated to outperform a uniform prior (Matzke and Wright, 2016). Other models, such as the Yule process or simple birth–death process are strongly violated when making phylogenetic inferences from paleontologic data. Moreover, the FBD tree prior has a high level of internal consistency for estimating age dates and a good fit to morphologic and geologic data in empirical, well-characterized data sets (e.g., penguins and canids, see Drummond and Stadler, 2016). Analyses of the crinoid data set herein assumed the simple case of constant rates for macroevolutionary and sampling parameters. However, the FBD process can be extended to a more sophisticated time-varying (i.e., piecewise-constant) model that may be useful for other data sets. Similarly, models could be developed to account for geographic variation in sampling probabilities (Wagner and Marcot, 2013). Such models may be especially beneficial for studies with larger, more comprehensive character matrices spanning similar to longer time intervals than considered herein (Gavryushkina et al., 2014; Zhang et al., 2016).

A major innovation of the FBD process as a tree prior is the ability to account for sampling ancestor–descendant relationships in phylogenetic analysis. Although the notion of discovering ‘true’ ancestors is somewhat contentious (see Smith, 1994; Foote, 1996), modeling studies suggest ancestral morphotaxa are likely present in the fossil record of many paleontologically important groups (Foote, 1996). Gavryushkina et al. (2014) demonstrated that sampled ancestors should be accounted for when estimating phylogenies, even when ancestral morphotaxa are not of specific interest in the analysis, because not including them introduces biases in parameter estimation. Thus, even if sampled ancestors are considered nuisance parameters (Close et al., 2015; Gorscak and O’Connor, 2016), they may nevertheless be important for accurately estimating more accurate tree topologies and node ages.

Parameter estimation under the FBD process does not require exhaustive sampling of fossil taxa, but it does require a representative random sample of species (e.g., the sampling strategy used herein) (Didier et al., 2012). However, the application of sampled-ancestor tip-dating methods to exhaustively sampled species-level (sensu Smith, 1994) or specimen-level data represents an important (but unexplored) frontier in phylogeny-based analyses of macroevolution. For example, coding multiple fossil specimens of species-level morphotaxa from different time horizons and/or geographic localities may provide a means for testing whether speciation events occur primarily through budding or bifurcating cladogenesis (Gavryushkina et al., 2014; Hunt and Slater, 2016). In addition,

paleontologists are commonly interested in whether morphologic change is punctuated or gradual (Eldredge and Gould, 1972; Hunt, 2008). Testing alternative probabilistic models of character evolution within a similar phylogenetic and sampling framework as described above, where each model makes different assumptions about punctuated versus gradual rates of change (as well as durations of stasis), would provide additional insight into the dynamics of morphologic evolution during speciation events (see Wagner and Marcot, 2010).

The emerging synthesis between paleontology and model-based phylogenetics contributes to the growing consensus that research programs in systematic paleontology are greatly enhanced when grounded in rigorous analytical approaches (Smith, 1994; Wagner, 2000b; Wagner and Marcot, 2010; Slater and Harmon, 2013; Hunt and Slater, 2016). Many of the analytical tools discussed in this paper were originally developed for non-paleontologic purposes. Thus, it is perhaps not surprising there is plenty of room for future modification and refinement of these techniques to better test paleontologic patterns. Nevertheless, the development and application of these methods has already expanded our ability to quantitatively address macroevolutionary questions. Continued research implementing probabilistic approaches in phylogeny-based paleontology will likely return the favor and provide neontologists with evolutionary insights and unique perspectives only accessible to paleontologists.

## Acknowledgments

This paper stems from research conducted at The Ohio State University toward the completion of a PhD in Geological Sciences. I thank W.I. Ausich, S.R. Cole, and T.W. Kammer for numerous discussions on crinoid morphology and evolution that helped shape some of the ideas in this manuscript. A. Gavryushkina is thanked for generously providing feedback on mathematical aspects of this paper. G.J. Slater and an anonymous reviewer are thanked for reviewing the manuscript. In particular, I thank G.J. Slater for helpful comments and suggestions regarding sensitivity analyses. S.R. Cole provided helpful assistance illustrating crinoids. K. Hollis, T. Ewin, P. Mayer, and K. Riddigton are thanked for providing access to museum specimens. Lastly, I especially thank S. Edie for kindly providing me with a place to stay while visiting the Field Museum collections. This research was supported by numerous student research grants from the Paleontological Society, Sigma Xi, the Palaeontological Association, the American Federation of Mineralogical and Geological Societies, Friends of Orton Hall, and a Presidential Fellowship from The Ohio State University, as well as a National Science Foundation’s Assembling the Echinoderm Tree of Life grant (DEB 1036416) to W. I. Ausich.

## Accessibility of supplemental data

Data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.6hb7j>

## References

- Angelin, N.P., 1878, Iconographia Crinoideorum. in *Stratis Sueciae Siluricis fossilium*: Holmia, Samson and Wallin, 62 p.
- Alroy, J., 2010, Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification: *Palaeontology*, v. 53, p. 1211–1235.
- Ausich, W.I., 1998, Phylogeny of Arenig to Caradoc Crinoids (Phylum Echinodermata) and suprageneric classification of the Crinoidea: The University of Kansas Paleontological Contributions Papers, New Series, v. no. 9, 36 p.
- Ausich, W.I., and Kammer, T.W., 2001, The study of crinoids during the 20th century and the challenges of the 21st century: *Journal of Paleontology*, v. 75, p. 1161–1173.
- Ausich, W.I., Kammer, T.W., and Baumiller, T.K., 1994, Demise of the middle Paleozoic crinoid fauna: A single extinction event or rapid faunal turnover?: *Paleobiology*, v. 20, p. 345–361.
- Ausich, W.I., Kammer, T.W., Rhenberg, E.C., and Wright, D.F., 2015, Early phylogeny of crinoids within the Pelmatozoan clade: *Palaeontology*, v. 58, p. 937–952.
- Bapst, D.W., 2012, When can clades be potentially resolved with morphology?: *PLoS One*, v. 8, e62312 doi: 10.1371/journal.pone.0062312.
- Bapst, D.W., 2013, A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa: *Methods in Ecology and Evolution*, v. 4, p. 724–733.
- Bapst, D.W., and Hopkins, M.J., in press, Comparing *cal3* and other *a posteriori* time-scaling approaches in a case study with the Pteroccephaliid trilobites: *Paleobiology*.
- Bapst, D.W., Wright, A.M., Matzke, N.J., and Lloyd, G.T., 2016, Topology, divergence dates, and macroevolutionary inferences vary between different tip-dating approaches applied to fossil theropods (Dinosauria): *Biology Letters*, v. 12, 20160237, doi: 10.1098/rsbl.2016.0237.
- Bather, F.A., 1890, British fossil crinoids. I. Historical introduction: *Annals and Magazine of Natural History*, ser. 6, v. 5, p. 306–310.
- Bather, F.A., 1899, A phylogenetic classification of the Pelmatozoa: *British Association for the Advancement of Science*, v. 1898, p. 916–923.
- Bather, F.A., 1900, Part III The Echinodermata. The Pelmatozoa, in *Lankester, E.R., ed., A Treatise on Zoology*: London, Adam and Charles Black, p. 94–204.
- Bell, M.A., and Lloyd, G.T., 2015, strap: An R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence: *Palaeontology*, v. 58, p. 379–389.
- Bergsten, J., Nilsson, A.N., and Ronquist, F., 2013, Bayesian tests of topology hypotheses with an example from diving beetles: *Systematic Biology*, v. 62, p. 660–673.
- Billings, E., 1857, New species of fossils from Silurian rocks of Canada. Canada Geological Survey, Report of Progress 1853–1856: Report for the year, v. 1856, p. 247–345.
- Billings, W.R., 1887, A new genus and three new species of crinoids from the Trenton Formation with notes on a large specimen of *Dendrocrinus proboscidiatus*: The Ottawa Naturalist (Ottawa Field Naturalists' Club Transactions), v. 111, p. 49–54.
- Bokma, F., 2008, Detection of “punctuated equilibrium” by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny: *Evolution*, v. 62, p. 2718–2726.
- Brower, J.C., 1995, Dendrocrinid crinoids from the Ordovician of northern Iowa and southern Minnesota: *Journal of Paleontology*, v. 69, p. 939–960.
- Brusatte, S.L., 2010, Representing supraspecific taxa in higher-level phylogenetic analyses: Guidelines for palaeontologists: *Palaeontology*, v. 53, p. 1–9.
- Brusatte, S.L., Benton, M.J., Ruta, M., and Lloyd, G.T., 2008, Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs: *Science*, v. 321, p. 1485–1488.
- Carlson, S.J., and Fitzgerald, P.C., 2007, Sampling taxa, estimating phylogeny and inferring macroevolution: An example from Devonian terebratulide brachiopods: *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, v. 98, p. 311–325.
- Clarke, J.A., and 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: *Systematic Biology*, v. 57, p. 185–201.
- Cole, S.R., 2017, Phylogeny and morphologic evolution of the Ordovician Camerata (Class Crinoidea, Phylum Echinodermata): *Journal of Paleontology*, doi:10.1017/jpa.2016.137.
- Conrad, T.A., 1842, Descriptions of new species of organic remains belonging to the Silurian, Devonian and Carboniferous Systems of the United States: *Philadelphia Journal of the Academy Natural Sciences of Philadelphia*, ser. 1, v. 8, no. 2, p. 228–280.
- Close, R.A., Friedman, M., Lloyd, G.T., and Benson, R.B., 2015, Evidence for a mid-Jurassic adaptive radiation in mammals: *Current Biology*, v. 25, p. 2137–2142.
- Deline, B., and Ausich, W.I., 2011, Testing the plateau: A reexamination of disparity and morphologic constraints in early Paleozoic crinoids: *Paleobiology*, v. 37, p. 214–236.
- Didier, G., Royer-Carenzi, M., and Laurin, M., 2012, The reconstructed evolutionary process with the fossil record: *Journal of Theoretical Biology*, v. 315, p. 26–37.
- Donoghue, P.C.J., and Benton, M.J., 2007, Rocks and clocks: Calibrating the Tree of Life using fossils and molecules: *Trends in Ecology and Evolution*, v. 22, p. 424–431.
- dos Reis, M., Donoghue, P.C., and Yang, Z., 2016, Bayesian molecular clock dating of species divergences in the genomics era: *Nature Reviews Genetics*, v. 17, p. 71–80.
- Drummond, A.J., and Rambaut, A., 2007, BEAST: Bayesian evolutionary analysis by sampling trees: *BMC Evolutionary Biology*, v. 7, doi: 10.1186/1471-2148-7-214.
- Drummond, A. J., and Stadler, T., 2016, Bayesian phylogenetic estimation of fossil ages: arXiv, preprint, 1601.07447v1.
- Eldredge, N., and Gould, S.J., 1972, Punctuated equilibria: An alternative to phyletic gradualism, in *Schopf, T.J.M., ed., Models in Paleobiology*: San Francisco, Freeman, Cooper and Company, p. 82–115.
- Feller, W., 1968, *An Introduction to Probability Theory and its Applications*, 1 (third edition): New York, Wiley, 509 p.
- Felsenstein, J., 1985, Phylogenies and the comparative method: *American Naturalist*, v. 126, p. 1–15.
- Felsenstein, J., 2004, *Inferring phylogenies*, volume 2: Sunderland, Sinauer Associates, 664 p.
- Foot, M., 1994, Morphological disparity in Ordovician-Devonian crinoids and the early saturation of morphological space: *Paleobiology*, v. 20, p. 320–344.
- Foot, M., 1996, On the probability of ancestors in the fossil record: *Paleobiology*, v. 22, p. 141–151.
- Foot, M., 1997, Estimating taxonomic durations and preservation probability: *Paleobiology*, v. 23, p. 278–300.
- Foot, M., 2000, Origination and extinction components of taxonomic diversity: General problems: *Paleobiology*, v. 26, p. 74–102.
- Foot, M., and Raup, D.M., 1996, Fossil preservation and the stratigraphic ranges of taxa: *Paleobiology*, v. 22, p. 121–140.
- Frest, T.J., and Strimple, H.L., 1978, *Manicrinus* (Nov.), a cladid evolutionary homeomorph of the bottom dwelling *Hybocrinus*, Brownsport (Silurian, Ludlow) of Tennessee: *Southeastern Geology*, v. 19, p. 157–175.
- Gahn, F.J., and Kammer, T.W., 2002, The cladid crinoid *Barycrinus* from the Burlington Limestone (early Osagean) and the phylogenetics of Mississippian botryocrinids: *Journal of Paleontology*, v. 76, p. 123–133.
- Gavryshkina, A., Welch, D., Stadler, T., and Drummond, A. J., 2014, Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration: *PLoS Computational Biology*, v. 10, e1003919.
- Gavryshkina, A., Heath, T.A., Ksepka, D.T., Stadler, T., Welch, D., and Drummond, A.J., 2015, Bayesian total evidence dating reveals the recent crown radiation of penguins: arXiv, preprint, v. 1506, 04797.
- Gelman, A., and Rubin, D.B., 1992, Inference from iterative simulation using multiple sequences: *Statistical Science*, v. 7, p. 457–472.
- Gil Cid, M.D., Alonso, P.D. and Pobes, E.S., 1996, Reconstrucción y modo de vida de *Heviacrinus melendezi* nov. gen. nov. sp. (Disparida, Iocrinidae), primer crinoide descrito del Ordovícico medio de los Montes de Toledo (España): *Revista de la Sociedad Geológica España*, v. 9, no. 1–2, p. 19–27.
- Goldfuss, G.A., 1826–1844, *Petrefacta Germaniae*, tam ea, Quae in Museo Universitatis Regiae Borussiae Fridericiae Wilhelmae Rhenanae, servantur, quam alia quaecunque in Museis Hoeninghusiano Muensteriano aliisque, extant, iconibus et descriptionibus illustrata. — Abbildungen und Beschreibungen der Petrefacten Deutschlands und der Angränzende Länder, unter Mitwirkung des Herrn Grafen Georg zu Münster, herausgegeben von August Goldfuss. v. 1 (1826–1833), Divisio prima. Zoophytorum reliquiae, p. 1–114; Divisio secunda. Radiariorum reliquiae, p. 115–221 [Echinodermata]; Divisio tertia. Annulorum reliquiae, p. 222–242; v. 2 (1834–1840), Divisio quarta. Molluscorum acephalicorum reliquiae. I. Bivalvia, p. 65–286; II. Brachiopoda, p. 287–303; III. (1841–1844), Divisio quinta. Molluscorum gasteropodum reliquiae, p. 1–121: Düsseldorf, Arnz & Co.
- Goldring, W., 1923, *The Devonian crinoids of the state of New York*: New York State Museum, Memoir, v. 16, p. 1–670.
- Gorscak, E., and O'Connor, P.M., 2016, Time-calibrated models support congruency between Cretaceous continental rifting and titanosaurian evolutionary history: *Biology Letters*, v. 12, 20151047.
- Guensburg, T.E., 2010, *Alphacrinus* new genus and origin of the disparid clade: *Journal of Paleontology*, v. 84, p. 1209–1216.
- Guensburg, T.E., 2012, Phylogenetic implications of the oldest crinoids: *Journal of Paleontology*, v. 86, p. 455–461.
- Guensburg, T.E., and Sprinkle, J., 2003, The oldest known crinoids (Early Ordovician, Utah) and a new crinoid plate homology system: *Bulletins of American Paleontology*, v. 364, 43 p.
- Guensburg, T.E., and Sprinkle, J., 2009, Solving the mystery of crinoid ancestry: New fossil evidence of arm origin and development: *Journal of Paleontology*, v. 83, p. 350–364.

- Guillerme, T., and Cooper, N., 2016, Effects of missing data on topological inference using a total evidence approach: *Molecular Phylogenetics and Evolution*, v. 94, p. 146–158.
- Hall, J., 1852, *Palaeontology of New York*, Volume 2, Containing Descriptions of the Organic Remains of the Lower Middle Division of the New-York System. Natural History of New York, Part 6: New York, D. Appleton & Co. and Wiley & Putnam; Boston, Gould, Kendall, & Lincoln, 362 p.
- Harmon, L.J., Losos, J.B., Davies, T.J., Gillespie, R.G., Gittleman, J.L., *et al.* 2010, Early bursts of body size and shape evolution are rare in comparative data: *Evolution*, v. 64, p. 2385–2396.
- Harrison, L.B., and Larsson, H.C.E., 2015, Among-character rate variation distributions in phylogenetic analysis of discrete morphologic characters: *Systematic Biology*, v. 64, p. 307–324.
- Heath, T.A., and Moore, B.R., 2014, Bayesian inference of species divergence times, in Chen, M.-H., Kuo, L., and Lewis, P.O., eds., *Bayesian Phylogenetics: Methods, Algorithms, and Applications*: CRC Press, Oxfordshire, p. 277–318.
- Heath, T.A., Hedtke, S.M., and Hillis, D.M., 2008, Taxon sampling and the accuracy of phylogenetic analyses: *Journal of Systematics and Evolution*, v. 46, p. 239–257.
- Heath, T.A., Huelsenbeck, J.P., and Stadler, T., 2014, The fossilized birth–death process for coherent calibration of divergence-time estimates: *Proceedings of the National Academy of Sciences*, v. 111, p. E2957–E2966.
- Heled, J., and Bouckaert, R.R., 2013, Looking for trees in the forest: Summary tree from posterior samples: *BMC Evolutionary Biology*, v. 13, p. 221.
- Hemery, L.G., Roux, M., Ameziame, N., and Eleaume, M., 2013, High-resolution crinoid phyletic inter-relationships derived from molecular data: *Cahiers De Biologie Marine*, v. 54, p. 511–523.
- Hopkins, M.J., and Smith, A.B., 2015, Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution: *Proceedings of the National Academy of Sciences*, v. 112, p. 3758–3763.
- Huelsenbeck, J.P., Larget, B., Miller, R.E., and Ronquist, F., 2002, Potential applications and pitfalls of Bayesian inference of phylogeny: *Systematic Biology*, v. 51, p. 673–688.
- Huelsenbeck, J.P., Ronquist, F., and Teslenko, M., 2015, Command Reference for MrBayes ver. 3.2.5: <http://mrbayes.sourceforge.net/manual.php> (accessed June 2016).
- Hunt, G., 2008, Gradual or pulsed evolution: When should punctuational explanations be preferred?: *Paleobiology*, v. 34, p. 360–377.
- Hunt, G., and Carrano, M.T., 2010, Models and methods for analyzing phenotypic evolution in lineages and clades, in Alroy, J., and Hunt, G., eds., *Short Course on Quantitative Methods in Paleobiology*: New Haven, Connecticut, Paleontological Society, p. 245–269.
- Hunt, G., and Slater, G., 2016, Integrating paleontological and phylogenetic approaches to macroevolution: *Annual Review of Ecology, Evolution, and Systematics*, v. 47, p. 189–213.
- Hunt, G., Bell, M.A., and Travis, M.P., 2008, Evolution toward a new adaptive optimum: Phenotypic evolution in a fossil stickleback lineage: *Evolution*, v. 62, p. 700–710.
- Jablonski, D., 2008, Biotic interactions and macroevolution: Extensions and mismatches across scales and levels: *Evolution*, v. 62, p. 715–739.
- Jaekel, O., 1902, Über verschiedene Wege phylogenetischer Entwicklung: 5th Verhandlungen der International Zoological-Congress Berlin, v. 1901, p. 1058–1117.
- Jaekel, O., 1906, Der oberste Lenneschiefer zwischen Letmathe und Iserlohn, in Schmidt, W.E., *Zeitschrift der Deutschen Geologischen Gesellschaft*, v. 57, p. 544.
- Kammer, T.W., 2001, Phenotypic bradytely in the *Costalocrinus-Barycrinus* lineage of Paleozoic cladid crinoids: *Journal of Paleontology*, v. 75, p. 383–389.
- Kammer, T.W., and Ausich, W.I., 1992, Advanced cladid crinoids from the middle Mississippian of the east-central United States: Primitive-grade calyces: *Journal of Paleontology*, v. 66, p. 461–480.
- Kammer, T.W., and Ausich, W.I., 1996, Primitive cladid crinoids from upper Osagean-lower Meramecian (Mississippian) rocks of east-central United States: *Journal of Paleontology*, v. 70, p. 835–866.
- Kass, R.E., and Raftery, A.E., 1995, Bayes factors: *Journal of the American Statistical Association*, v. 90, p. 773–795.
- Kier, P.M., 1952, Echinoderms of the Middle Devonian Silica Formation of Ohio: University of Michigan Contributions from Museum of Paleontology, v. 10, p. 59–81.
- Koenig, J.W., and Meyer, D.L., 1965, Two new crinoids from the Devonian of New York: *Journal of Paleontology*, v. 39, p. 391–397.
- Ksepka, D.T., Parham, J.F., Allman, J.F., Benton, M.J., Carrano, M.T., Cranston, K.A., Donoghue, P.C., Head, J.J., Hermesen, E.J., Irmis, R.B., and Joyce, W.G., 2015, The Fossil Calibration Database—A new resource for divergence dating: *Systematic Biology*, v. 64, p. 853–859.
- Lakner, C., van der Mark, P., Huelsenbeck, J.P., Larget, B., and Ronquist, F., 2008, Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics: *Systematic Biology*, v. 57, p. 86–106.
- Lane, N.G., 1970, Lower and Middle Ordovician crinoids from west-central Utah: Brigham Young University Geology Studies, v. 17, p. 3–17.
- Lee, M.S.Y., and Palci, A., 2015, Morphological phylogenetics in the genomic age: *Current Biology*, v. 25, p. R922–R929.
- Lee, M.S., Cau, A., Naish, D., and Dyke, G.J., 2014, Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves: *Systematic Biology*, v. 63, p. 442–449.
- Lepage, T., Bryant, D., Philippe, H., and Lartillot, N., 2007, A general comparison of relaxed molecular clock models: *Molecular Biology and Evolution*, v. 24, p. 2669–2680.
- Lewis, P.O., 2001, A likelihood approach to estimating phylogeny from discrete morphological character data: *Systematic Biology*, v. 50, p. 913–925.
- Lloyd, G.T., Wang, S.C., and Brusatte, S.L., 2012, Identifying heterogeneity in rates of morphological evolution: Discrete character change in the evolution of lungfish (Sarcopterygii: Dipnoi): *Evolution*, v. 66, p. 330–348.
- Maddison, D.R., 1991, The discovery and importance of multiple islands of most-parsimonious trees: *Systematic Zoology*, v. 40, p. 315–328.
- Matzke, N.J., 2015, The evolution of antievolution policies after Kitzmiller v. Dover: *Science*, v. 351, p. 28–30, doi: 10.1126/science.aad4057.
- Matzke, N.J., and Wright, A., 2016, Ground truthing tip-dating methods using fossil Canidae reveals major differences in performance: *Biology Letters*, v. 12, 2010328, doi: 10.1098/rsbl.2016.0328.
- McIntosh, G.C., 1986, Phylogeny of the dicyclid inadunate order Cladida: Fourth North American Paleontological Convention Abstracts: University of Colorado, Boulder, p. A31.
- McIntosh, G.C., 2001, Devonian cladid crinoids: Families Glossocrinidae Goldring, 1923, and Rutkowskiicrinidae new family: *Journal of Paleontology*, v. 75, p. 783–807.
- Meek, F.B., 1871, On some new Silurian (Ordovician) crinoids and shells: *American Journal of Science*, ser. 3, v. 1, p. 295–299.
- Meek, F.B., and Worthen, A.H., 1865, Descriptions of new species of crinoidea, etc. from the Paleozoic rocks of Illinois and some of the adjoining states: *Proceedings of the Academy of Natural Sciences of Philadelphia*, v. 17, p. 143–155.
- Miller, S.A., and Gurley, W.F.E., 1895, New and interesting species of Palaeozoic fossils: Illinois State Museum, Bulletin, v. 7, 89 p.
- Moore, R.C., and Laudon, L.R., 1943, Evolution and classification of Paleozoic crinoids: Geological Society of America Special Paper, v. no. 46, 151 p.
- Moore, R.C., and Teichert, C., eds., 1978, *Treatise on Invertebrate Paleontology*, Part T, Echinodermata 2: Lawrence, Kansas, Geological Society of America and University of Kansas Press, 1027 p.
- Müller, J., 1859, Bericht über die zur Bekanntmachung geeigneten Verhandlungen der Königlich Preussischen Akademie der Wissenschaften zu Berlin: Monatsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin (1858), p. 185–198.
- O'Meara, B.C., Ané, C., Sanderson, M.J., Wainwright, P.C., and Hansen, T., 2006, Testing for different rates of continuous trait evolution using likelihood: *Evolution*, v. 60, p. 922–933.
- O'Reilly, J.E., Puttick, M.N., Parry, L., Tanner, A.R., Tarver, J.E., Fleming, J., Pisani, D., and Donoghue, P.C.J., 2016, Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data: *Biology Letters*, v. 12, 20160081, <http://dx.doi.org/10.1098/rsbl.2016.0081>.
- Paradis, E., Claude, J., and Strimmer, K., 2004, APE: Analyses of phylogenetics and evolution in R language: *Bioinformatics*, v. 20, p. 289–290.
- Pennell, M.W., Harmon, L.J., and Uyeda, J.C., 2014, Is there room for punctuated equilibrium in macroevolution?: *Trends in Ecology and Evolution*, v. 29, p. 23–32.
- Phillips, J., 1839, Encrinites and zoophytes of the Silurian System, in Murchison, R.T., ed., *The Silurian System*, p. 670–675.
- Pollitt, J.R., Fortey, R.A., and Wills, M.A., 2005, Systematics of the trilobite families Lichidae Hawle and Corda, 1847 and Lichakephalidae Tripp, 1957: The application of Bayesian inference to morphological data: *Journal of Systematic Palaeontology*, v. 3, p. 225–241.
- Pyrn, R.A., 2011, Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia: *Systematic Biology*, v. 60, p. 466–481.
- Pyrn, R.A., 2015, Post-molecular systematics and the future of phylogenetics: *Trends in Ecology and Evolution*, v. 30, p. 384–389.
- Rabosky, D.L., 2009, Ecological limits and diversification rate: Alternative paradigms to explain the variation in species richness among clades and regions: *Ecology Letters*, v. 12, p. 735–743.
- Rabosky, D.L., and McCune, A.R., 2009, Reinventing species selection with molecular phylogenies: *Trends in Ecology and Evolution*, v. 25, p. 68–74.
- Rambaut, A., 2014, Summarizing posterior trees: <http://beast.bio.ed.ac.uk/summarizing-posterior-trees> (accessed June 2016).

- Rambaut, A., and Drummond, A.J., 2015, TreeAnnotator v2 2.1: MCMC output analysis, <http://beast.bio.ed.ac.uk/treeannotator>.
- Rannala, B., and Yang, Z., 1996, Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference: *Journal of Molecular Evolution*, v. 43, p. 304–311.
- Raup, D.M., 1985, Mathematical models of cladogenesis: *Paleobiology*, v. 11, p. 42–52.
- Raup, D.M., Gould, S.J., Schopf, T.J.M., and Simberloff, D.S., 1973, Stochastic models of phylogeny and the evolution of diversity: *Journal of Geology*, v. 81, p. 525–542.
- Robinson, D.F., and Foulds, L.R., 1981, Comparison of phylogenetic trees: *Mathematical Biosciences*, v. 53, p. 131–147.
- Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L., and Rasnitsyn, A.P., 2012, A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera: *Systematic Biology*, v. 61, p. 973–999.
- Rouse, G.W., Jermiin, L.S., Wilson, N.G., Eeckhaut, I., Lanterbecq, D., Oji, T., Young, C.M., Browning, T., Cisternas, P., Helgen, L.E., Stuckey, M., and Messing, C.G., 2013, Fixed, free, and fixed: The fickle phylogeny of extant Crinoidea (Echinodermata) and their Permian–Triassic origin: *Molecular Phylogenetics and Evolution*, v. 66, p. 161–181.
- Salter, J.W., 1873, A Catalogue of the Collection of Cambrian and Silurian Fossils Contained in the Geological Museum of the University of Cambridge: Cambridge, Cambridge University Press, 204 p.
- Schlothem, E.F. von, 1820, Die Petrefactenkunde auf ihrem jetzigen Standpunkte durch die Beschreibung seiner Sammlung versteinerter und fossiler Überreste des Thier- und Pflanzenreichs der Vorwelt erläutert: Gotha, Beckersche Buchhandlung, 437 p.
- Schultze, L., 1867, Monographie der Echinodermen des Eifler Kalkes: Denkschriften der Kaiserlich Akademie der Wissenschaften Mathematisch-Naturwissenschaftlichen Classe, Wien, v. 26, no. 2, p. 113–230.
- Sepkoski, J.J., Jr., 1981, A factor analytic description of the Phanerozoic marine fossil record: *Paleobiology*, v. 7, p. 36–53.
- Sevastopulo, G.D., and Lane, N.G., 1988, Ontogeny and phylogeny of disparid crinoids, in Paul, C.R.C., and Smith, A.B., eds., *Echinoderm Phylogeny and Evolutionary Biology*: Oxford, Clarendon Press, p. 245–253.
- Simms, M.J., 1988, The phylogeny of post-Palaeozoic crinoids, in Burke, R.D., Mladenov, P.V., Lambert, P., and Parsley, R.L., eds., *Echinoderm Biology*: Rotterdam, Balkema, p. 97–102.
- Simms, M.J., and Sevastopulo, G.D., 1993, The origin of articulate crinoids: *Palaeontology*, v. 36, p. 91–109.
- Simpson, G.G., 1944, *Tempo and Mode in Evolution*: Columbia University Press, New York, 237 p.
- Slater, G.J., 2013, Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous–Palaeogene boundary: *Methods in Ecology and Evolution*, v. 4, p. 734–744.
- Slater, G.J., 2015, Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution: *Proceedings of the National Academy of Sciences*, v. 112, p. 4897–4902.
- Slater, G.J., and Harmon, L.J., 2013, Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution: *Methods in Ecology and Evolution*, v. 4, p. 699–702.
- Smith, A.B., 1984, Classification of the Echinodermata: *Palaeontology*, v. 27, p. 431–459.
- Smith, A.B., 1994, *Systematics and the Fossil Record: Documenting Evolutionary Patterns*: Oxford, Blackwell Science, 223 p.
- Smith, A.B., Lafay, B., and Christen, R., 1992, Comparative variation of morphological and molecular evolution through geologic time: 28S ribosomal RNA versus morphology in echinoids: *Philosophical Transactions: Biological Sciences*, v. 338, p. 365–382.
- Snively, E., Russell, A.P., and Powell, G.L., 2004, Evolutionary morphology of the coelurosaurian arctometatarsus: Descriptive, morphometric and phylogenetic approaches: *Zoological Journal of the Linnean Society*, v. 142, p. 525–553.
- Spencer, M.R., and Wilberg, E.W., 2013, Efficacy or convenience? Model-based approaches to phylogeny estimation using morphological data: *Cladistics*, v. 29, p. 663–671.
- Springer, F., 1911, On a Trenton echinoderm fauna: Canada Department Mines Memoir no. 15-P, 70 p.
- Springer, F., 1920, *The Crinoidea Flexibilia*: Smithsonian Institution Publication no. 2501, 486 p.
- Sprinkle, J., 1982, Hybocrinus, in Sprinkle, J., ed., *Echinoderm Faunas from the Bromide Formation (Middle Ordovician) of Oklahoma*: The University of Kansas Paleontological Contributions, Monograph, v. 1, p. 119–128.
- Sprinkle, J., and Kolata, D.R., 1982, “Rhomb-bearing” camerate, in Sprinkle, J., ed., *Echinoderm Faunas from the Bromide Formation (Middle Ordovician) of Oklahoma*: University of Kansas Paleontological Contributions, Monograph, v. 1, p. 206–211.
- Sprinkle, J., and Moore, R.C., 1978, Hybocrinida, in Moore, R. C., and Teichert, C., eds., *Treatise on Invertebrate Paleontology, Part T, Echinodermata 2*: Lawrence, Kansas, Geological Society of America and University of Kansas Press, p. T564–T574.
- Stadler, T., 2010, Sampling-through-time in birth-death trees: *Journal of Theoretical Biology*, v. vol. 267, p. 396–404.
- Stadler, T., and Yang, Z., 2013, Dating phylogenies with sequentially sampled tips: *Systematic Biology*, v. 62, p. 674–688.
- Stadler, T., Kouyos, R., Wyl, V., von, Yerly, S., Böni, J., Bürgisser, P., Klimkait, T., Joos, B., Rieder, P., Xie, D., Günthard, H.F., Drummond, A.J., and Bonhoeffer, S., the Swiss HIV Cohort Study, 2012, Estimating the basic reproductive number from viral sequence data: *Molecular Biology and Evolution*, v. 29, p. 347–357.
- Summers, M.M., Messing, C.G., and Rouse, G.W., 2014, Phylogeny of Comatulidae (Echinodermata: Crinoidea: Comatulida): A new classification and an assessment of morphological characters for crinoid taxonomy: *Molecular Phylogenetics and Evolution*, v. 80, p. 319–339.
- Swofford, D.L., 2002, PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods): Version 4: Massachusetts, Sinauer Associates.
- Swofford, D.L., Olsen, G.J., Waddell, P.J., and Hillis, D.M., 1996, Phylogenetic inference, in Hillis, D.M., Moritz, C., and Mable, B.K., eds., *Molecular Systematics* (second edition): Massachusetts, Sinauer, p. 407–514.
- Ubahgs, G., 1969, *Aethocrinus moorei* Ubahgs n. gen., n. sp., le Plus ancien Crinoïde dicyclique Connus: University of Kansas Paleontological Contributions, Paper 38, p. 1–25.
- Ubahgs, G., 1978, Origin of crinoids, in Moore, R.C., and Teichert, C., eds., *Treatise on Invertebrate Paleontology, Part T, Echinodermata 2*: Lawrence, Geological Society of America and University of Kansas Press, p. T275–T281.
- Ulrich, E.O., 1925, The lead, zinc, and fluorspar deposits of Western Kentucky: Chapter 2, Stratigraphic geology: U. S. Geological Survey, Professional Paper 36, p. 22–71.
- Wagner, P.J., 1998, A likelihood approach for evaluating estimates of phylogenetic relationships among fossil taxa: *Paleobiology*, v. 24, p. 430–449.
- Wagner, P.J., 1999, The utility of fossil data in phylogenetic analyses: A likelihood example using Ordovician–Silurian species of the Lophospiridae (Gastropoda: Murchisoniina): *American Malacological Bulletin*, v. 15, p. 1–32.
- Wagner, P.J., 2000a, Phylogenetic analyses and the fossil record: Tests and inferences, hypotheses, and models: *Paleobiology*, v. 26, p. 341–371.
- Wagner, P.J., 2000b, The quality of the fossil record and the accuracy of phylogenetic inferences about sampling and diversity: *Systematic Biology*, v. 49, p. 65–86.
- Wagner, P.J., 2000c, Exhaustion of morphologic character states among fossil taxa: *Evolution*, v. 54, p. 365–386.
- Wagner, P.J., 2012, Modelling rate distributions using character compatibility: Implications for morphological evolution among fossil invertebrates: *Biology Letters*, v. 8, p. 143–146.
- Wagner, P.J., and Marcot, J.D., 2010, Probabilistic phylogenetic inference in the fossil record: Current and future applications, in Alroy, J., and Hunt, G., eds., *Quantitative Methods in Paleobiology: The Paleontological Society Papers*, v. 16, p. 189–211.
- Wagner, P.J., and Marcot, J.D., 2013, Modelling distributions of fossil sampling rates over time, space and taxa: Assessment and implications for macroevolutionary studies: *Methods in Ecology and Evolution*, v. 4, p. 703–713.
- Walcott, C.D., 1884, Descriptions of new species of fossils from the Trenton Group of New York: New York State Museum of Natural History, Annual Report, v. 35, p. 207–214.
- Webster, G.D., 2003, Bibliography and index of Paleozoic crinoids, coronates, and hemistreptocrinoids, 1758–1999: Geological Society of America Special Paper 363.
- Webster, G.D., and Jell, P.A., 1999, New Permian crinoids from Australia: *Memoirs of the Queensland Museum*, v. 33, p. 349–359.
- Webster, G.D., and Maples, C.G., 2006, Cladid crinoid (Echinodermata) anal conditions: A terminology problem and proposed solution: *Palaeontology*, v. 49, p. 187–212.
- Webster, G.D., and Maples, C.G., 2008, Cladid crinoid radial facets, brachials, and arm appendages: A terminology solution for studies of lineage, classification, and paleoenvironment, in Ausich, W.I., and Webster, G.D., eds., *Echinoderm Paleobiology*: Bloomington, Indiana University Press, p. 197–226.
- Webster, G.D., Maples, C.G., Mawson, R., and Dastanpour, M., 2003, A cladid-dominated Early Mississippian crinoid and conodont fauna from Kerman Province, Iran and revision of the glossocrinids and rhenocrinids: *Journal of Paleontology*, Memoir 60, v. 77, suppl. to no. 3, 35 p.
- Weller, S., and Davidson, A.D., 1896, *Petalocrinus mirabilis* (n. sp.) and a new American fauna: *Journal of Geology*, v. 4, p. 166–173.
- Wetherby, A.G., 1880, Remarks on the Trenton Limestone of Kentucky, with descriptions of new fossils from that formation and the Kaskaskia (Chester)

- Group, Sub-carboniferous: Journal of the Cincinnati Society of Natural History, v. 3, p. 144–160.
- Wheeler, W.C., 2012, Systematics: A course of lectures: Chichester, John Wiley and Sons, 426 p.
- Wheeler, W.C., and Pickett, K.M., 2007, Topology-Bayes versus Clade-Bayes in phylogenetic analysis: Molecular Biology and Evolution, v. 25, p. 447–453.
- Wright, A.M., and Hillis, D.M., 2014, Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data: PLoS One, v. 9, e109210.
- Wright, D.F., 2015, Fossils, homology, and “Phylogenetic Paleo-ontology”: A reassessment of primary posterior plate homologies among fossil and living crinoids with insights from developmental biology: Paleobiology, v. 41, p. 570–591.
- Wright, D.F., and Ausich, W.I., 2015, From the stem to the crown: Phylogeny and diversification of pan-cladid crinoids, in Zamora, S., and Rábano, I., eds., Progress in Echinoderm Paleobiology: Cuadernos del museo Geominero, 19, Instituto Geológico y Minero de España, p. 199–202.
- Wright, D.F., and Stigall, A.L., 2013, Phylogenetic revision of the Late Ordovician orthid brachiopod genera *Plaesiomys* and *Hebertella* from Laurentia: Journal of Paleontology, v. 87, p. 1107–1128.
- Wright, D.F., Ausich, W.I., Cole, S.R., Peter, M.E., and Rhenburg, E.C., 2017, Phylogenetic taxonomy and classification of the Crinoidea: Journal of Paleontology, doi:10.1017/jpa.2016.142.
- Xie, W., Lewis, P.O., Fan, Y., Kuo, L., and Chen, M.-H., 2010, Improving marginal likelihood estimation for Bayesian phylogenetic model selection: Systematic Biology, v. 60, p. 150–160.
- Yang, Z., 2014, Molecular Evolution: A Statistical Approach: Oxford, Oxford University Press, 492 p.
- Zhang, C., Stadler, T., Klopstein, S., Heath, T.A., and Ronquist, F., 2016, Total-evidence dating under the fossilized birth–death process: Systematic Biology, v. 65, p. 228–249.

Accepted 31 August 2016

## Appendix. Probability density of FBD phylogenetic trees

This appendix provides equations for calculating the probability density of a sampled ancestor phylogenetic tree generated by the fossilized birth–death process. These trees are used as Bayesian priors for phylogenies (i.e., topology and divergence times) in the tip-dating analysis presented in the main text. The equations presented here follow Stadler (2010), Didier et al. (2012), and Stadler et al. (2012), with subsequent modifications by Gavryushkina et al. (2014). I encourage readers to refer to these references for additional details on birth–death sampling models and their applications. Interested readers are also encouraged to see Gavryushkina et al. (2014) and Zhang et al. (2016) for a detailed description of a birth–death sampling process with time heterogeneous (i.e., piecewise-constant) rates, as well as Bapst’s (2013) probabilistic method to a posteriori timescale undated cladograms. Please note that the notation used in this paper differs in places from that of the references mentioned above to be more consistent with the paleontological literature (e.g., Foote, 1997, 2000; Wagner and Marcot, 2010; Bapst, 2012).

For simplicity, ordered trees are used in the following derivation. However, ordered trees are subsequently modified to labeled trees (using a conversion factor) to calculate likelihoods (Stadler, 2010). Ordered trees distinguish between left and right branches and identify nodes via a unique left–right path from the root, whereas labeled trees instead give sampled nodes distinct labels to describe branching patterns and divergences (Stadler, 2010; Gavryushkina et al., 2014). Each labeled phylogenetic tree ( $\Psi$ ) consists of two elements: a discrete ranked tree topology  $\Psi$  and a continuous time vector  $\tau$ . The ranked tree topology  $\Psi$  with ordered branching events is simply the branching pattern of common ancestry among  $m$  taxa distributed at the tips of the tree (i.e., an unscaled cladogram).

The time vector  $\tau$  assigns an unambiguous time to each node, such that it contains the following elements arranged in the exact descending temporal order in which they occur in the tree: the  $m - 1$  lineage splitting times ( $x_1 \dots x_{m-1}$ , where  $x_{m-1} < \dots < x_1$ ); the  $m$  tip-taxon sampling times ( $y_1 \dots y_m$ , where  $y_m < \dots < y_1$ ); and  $k$  sampling times of two-degree nodes ( $z_1 \dots z_k$ , where  $z_k < \dots < z_1$ ) (Gavryushkina et al., 2014). The density of an ordered tree can be converted to a labeled tree by multiplying by the conversion factor  $[2^{n+m-1}/n!(m+k)!]$  (see Stadler, 2010, p. 402).

To obtain the probability density of a given sampled ancestor phylogenetic tree ( $\Psi$ ), the likelihood is calculated along each branch  $b$  in  $\Psi$  moving backward in time. The probability of any one event (i.e., branching with rate  $p$ , extinction with rate  $q$ , or sampling with rate  $r$ ) occurring during a very small time step  $\Delta t$  is the product of its Poisson rate and  $\Delta t$ . For example, the probability of a single branching event over a small time interval is  $p\Delta t$ . The probabilities corresponding to any event happening more than once during  $\Delta t$  is summarized by an order term,  $O(\Delta t^2)$  (Feller, 1968). If  $\Delta t$  is chosen to be very small, then the probability of more than one event happening during the time interval can be ignored (see below). Because time is measured from the tips backward toward the root, each small time step  $\Delta t$  moves further into the past.

Let  $\Psi_b(t)$  be the probability density that a lineage corresponding to branch  $b$  at time  $t$  produced the observed  $\Psi$  between time  $t$  and  $t = 0$ . Thus,

$$\Psi_b(t_o) = f[\Psi \mid \pi = (p, q, r, \varepsilon, t_o)].$$

To obtain the probability density for  $\Psi_b(t)$ , I follow Stadler et al. (2012) and first consider describing  $\Psi_b(t + \Delta t)$  and assume  $\Psi_b(t)$  is known. After deriving an equation for how the probability density changes over a small time interval  $\Delta t$ , the definition of a derivative can be used to obtain a differential equation describing the change in probability toward the root. This differential equation is then integrated along branches to solve for  $\Psi_b(t)$  across the phylogeny.

Consider the possible evolutionary histories for single lineage along  $b$  from time  $t$  to  $\Delta t$  in the past. Moving down the branch, a lineage may have either undergone at least one branching event during  $\Delta t$  or experienced no event at all (note that extinction is not considered here because a lineage cannot logically have gone extinct prior to its subsequent sampling). Thus, the equation for  $\Psi_b(t + \Delta t)$  is

$$\Psi_b(t + \Delta t) = (1 - (p + q + r)\Delta t - O(\Delta t^2))\Psi_b(t) + p\Delta t 2S_0(t)\Psi_b(t) + O(\Delta t^2) \quad (A1),$$

where  $S_0(t)$  is the probability that a lineage at time  $t$  has no sampled descendants (Stadler, 2010). The quantity  $S_0(t)$  is given by Gavryushkina et al. (2014) as

$$S_0(t) = \frac{p + q + r + c_1 \frac{e^{-c_1(1-c_2)} - (1+c_2)}{e^{-c_1(1-c_2)} + (1+c_2)}}{2p},$$

where

$$c_1 = \left| \sqrt{(p + q + r)^2 + 4pr} \right| \text{ and } c_2 = \frac{p - q - 2p\varepsilon - r}{c_1}.$$

The logic behind Equation (A1) is straightforward to interpret if each term on the right-hand side is considered



in isolation. The first term,  $(1 - (p + q + r)\Delta t - O(\Delta t^2))$ , is the probability that no event takes place along  $b$  during  $\Delta t$ . The second term reflects the probability of lineage splitting, where  $p\Delta t$  is the probability that one branching event takes place and  $2S_0(t)$  accounts for the probability that one of the two descendant lineages (either the left or right descendant) left no sampled descendants in the observed tree. The final term,  $O(\Delta t^2)$ , is a summary term corresponding to the probabilities for multiple events during  $\Delta t$ .

Now that we have an expression for  $\Psi_b(t + \Delta t)$ , if we modify Equation (A1) to consider how the probability density changes from time  $t$  to time  $t + \Delta t$ , we get

$$\frac{\Psi_b(t + \Delta t) - \Psi_b(t)}{\Delta t} = -(p + q + r)\Psi_b(t) + 2pS_0(t)\Psi_b(t) + O(\Delta t).$$

Finally, taking the limit as  $\Delta t \rightarrow 0$  results in the following differential equation:

$$\frac{d}{dt}\Psi_b(t) = -(p + q + r)\Psi_b(t) + 2pS_0(t)\Psi_b(t). \quad (\text{A2})$$

Letting  $T_e$  represent the terminal end of a branch  $b$ , then the initial values at  $t = T_e$  are

$$\Psi_b(T_e) = \begin{cases} p\Psi_{b1}(T_e)\Psi_{b2}(T_e) & \text{if } b \text{ has two descendant branches, } b_1 \text{ and } b_2 \\ r\Psi_{b1}(T_e) & \text{if } b \text{ has one descendant branch } b_1 \\ rS_0(T_e) & \text{if } b \text{ has no descendant branches and } T_e > 0 \\ \varepsilon & \text{if } b \text{ has no descendant branches and } T_e = 0. \end{cases}$$

Conditioning on the clade's origin (i.e., the age of the root) and the event  $S_e$ , where  $S_e$  denotes that at least one lineage was sampled at the end of the process, the closed form solution for the probability density of  $\Psi$  accounting for labeled trees is (Stadler, 2010, p. 400; Gavryushkina et al., 2014, equation 3)

$$\Psi_b(t_o) = f[\Psi \mid \pi = (p, q, r, \varepsilon, t_o, S_e)] \\ = \frac{1}{(m+k!)} \frac{r^k \varepsilon^n \varphi(t_o)}{(p(1 - \hat{S}_0(t_o)))^2} \prod_{i=1}^{m+n-1} 2p\varphi(x_i) \prod_{i=1}^m \frac{rS_0(y_i)}{\varphi(y_i)}, \quad (\text{A3})$$

where  $n$  is the number of  $\varepsilon$  sampled lineages,  $S_0$ ,  $c_1$ , and  $c_2$  are defined as above, with

$$\varphi(x) = \frac{4}{2(1 - c_2^2) + e^{-c_1 x}(1 - c_2)^2 + e^{c_1 x}(1 + c_2)^2},$$

and

$$\hat{S}_0(t_o) = 1 - \frac{\varepsilon(p - q)}{p\varepsilon + (p(1 - \varepsilon) - q)e^{-(p - q)t}}.$$