



Extant-only comparative methods fail to recover the disparity preserved in the bird fossil record

Jonathan S. Mitchell^{1,2,3}

¹Committee on Evolutionary Biology, University of Chicago, 5034 S. Woodlawn, Chicago, Illinois, 60615

²Ecology and Evolutionary Biology Department, The University of Michigan, 2019 Kraus Nat. Sci. Bldg., 830 N. University, Ann Arbor, Michigan, 48109

³E-mail: mitchelljs@uchicago.edu

Received March 21, 2015

Accepted July 12, 2015

Most extant species are in clades with poor fossil records, and recent studies of comparative methods show they have low power to infer even highly simplified models of trait evolution without fossil data. Birds are a well-studied radiation, yet their early evolutionary patterns are still contentious. The fossil record suggests that birds underwent a rapid ecological radiation after the end-Cretaceous mass extinction, and several smaller, subsequent radiations. This hypothesized series of repeated radiations from fossil data is difficult to test using extant data alone. By uniting morphological and phylogenetic data on 604 extant genera of birds with morphological data on 58 species of extinct birds from 50 million years ago, the “halfway point” of avian evolution, I have been able to test how well extant-only methods predict the diversity of fossil forms. All extant-only methods underestimate the disparity, although the ratio of within- to between-clade disparity does suggest high early rates. The failure of standard models to predict high early disparity suggests that recent radiations are obscuring deep time patterns in the evolution of birds. Metrics from different models can be used in conjunction to provide more valuable insights than simply finding the model with the highest relative fit.

KEY WORDS: Adaptive radiation, fossils, macroevolution, morphological evolution, paleobiology.

The ecological and evolutionary processes that control how clades diversify in morphological space are complex (e.g., Foote 1997), and when combined with the impacts of sporadic extinction events and geographic changes through time, it is impossible to fully understand a clade’s history using data from only extant forms. Unfortunately, most clades have poor fossil records, and much of the history of life has been lost through the biases of the rock record, although there are particular clades with exceptional representation (Behrensmeyer et al. 2000). Incorporating fossil data can have a profound impact on inferences of evolutionary patterns, even when the record is incomplete (e.g., Finarelli and Flynn 2006; Slater et al. 2012; Slater 2013; Pant et al. 2014).

Despite these complexities, if evolutionary biologists are intent on documenting the history of life, we need methods that can at least approximate patterns of evolution in deep time for clades without fossil information. Methods based on extant taxa

are strongly limited by a near-universal requirement of monotonic dynamics through time (although see Morlon et al. 2011). Simulation-based studies have shown that many commonly used methods lack the power to discriminate between different models reliably (Boettiger et al. 2012; Slater and Pennell 2013), and the mismatch between the patterns informed by the extant-only comparative approaches and the patterns observed in the fossil record are stark. Analyses of extant-only datasets consistently find little-to-no signal of rapid early morphological diversification (e.g., Harmon et al. 2010), whereas analyses of fossil data regularly support a pattern of morphological evolution in which a disproportionate amount of morphological evolution in a clade happens in the first half of a clade’s history (e.g., Hughes et al. 2013).

Fossil data can improve our understanding of certain clades, as well as serve as a check on various methods that may be too

sensitive to recent divergences to infer ancient patterns. Ancient species can only be placed on phylogenies using morphological data, which makes integrating information from the fossil record with molecular phylogenies difficult. However, the phenotypic variance observed within fossil species can be known with relatively high precision, allowing us to understand how disparity has changed through time. By comparing different metrics of fit between trait distributions and molecular phylogenies (rate estimates, changes in rates, patterns of subclade disparity, etc.) to disparity trajectories observed in the fossil record, we can begin to understand not only where our simplified models are inadequate but also how to interpret conflicting signals.

Recent advances in our understanding of both avian phylogeny (Hackett et al. 2008; Jetz et al. 2012; Jarvis et al. 2014) and the completeness of the avian fossil record (Behrensmeyer et al. 2003; Turvey and Blackburn 2011; Ksepka and Boyd 2012; Mitchell 2015) have made birds an attractive system for understanding how different methods for inferring deep time evolution relate to one another. Given that we have only recently had access to high-resolution data on the phylogeny of birds, most of our understanding of the deep patterns of avian evolution has come exclusively from paleontology. Early interpretations of the bird fossil record all focused on the apparent high rates of evolution immediately after the end-Cretaceous mass extinction, and suggested that there was a burst of ecological innovation directly associated with that event (e.g., Feduccia 1995). Although early studies lacked a phylogenetic framework, and were based on a potentially too-literal reading of the fossil record, the general impression of rapid diversification has only gained further support. Mayr (2005) noted the incredible diversity of nonpasserine birds in the Paleogene, as essentially all extant ecologies are present in the Paleogene, including specialists such as nectar-feeding forms (Mayr and Wilde 2015), aerial insectivores (Ksepka et al. 2013), wading birds (Smith et al. 2013), and sallying frogmouths (Nesbitt et al. 2011). These ecological specialists all evolved by about 50 million years ago after the origin of crown Aves, and about 15 million years ago after the end-Cretaceous mass extinction, which wiped out the earlier radiations of flighted stem birds (Benson and Choiniere 2013; Brusatte et al. 2014; Mitchell and Makovicky 2014).

The wide array of ecological forms in the Eocene fossil record has lead evolutionary biologists to hypothesize an early burst in all avians, but also to hypothesize about a subsequent radiation in Passeriformes specifically (Mayr and Manegold 2004; Mayr 2005; Livezey and Zuis 2007; Manegold 2008; Barker 2011). This sets up a hypothesis of two major ecological radiations in crown birds: once at the base of the tree and another again with the origin of Passeriformes. The primary way of detecting adaptive radiations from phylogenies of living species is to assume that the rate of morphological evolution decays through time, which

means that a wholesale analysis of living birds is likely unable to test this dual-radiation hypothesis. Understanding how quickly birds diversified, and testing hypotheses of complex dynamics such as these, requires uniting fossil data with molecular phylogenies and morphological data from extant birds.

In this study, I build on previous phylogenetic work (e.g., Jetz et al. 2012) and studies of morphological evolution (e.g., Schweizer et al. 2014) in birds to explore how effectively the deep history of the avian ecological radiation can be inferred. The most recent estimate for the age of crown Aves is in the Cretaceous, approximately 100 million years ago (Jetz et al. 2012; Jarvis et al. 2014), and two exceptional fossil deposits (Lagerstätten) are known from the “halfway point” of avian evolution (about 50 million years ago): North America’s Green River Formation (52 million years ago) and Germany’s Messel pits (47 million years ago). I fit models of phenotypic evolution to phylogenies of extant taxa, and used the parameters from those models to predict the expected disparity at the “halfway point” of avian evolution, which I compared to the disparity observed in the fossil record. These deposits preserve a suite of complete fossil bird skeletons that are both amenable to morphometric analysis and that have been robustly placed into a phylogenetic framework (e.g., Mayr 2005; Grande 2013).

The deep divergence among avian lineages has received a lot of critical attention, and different datasets have yielded conflicting results (e.g., James 2005; Brown et al. 2008; Pacheco et al. 2011; Ksepka and Boyd 2012). Phenotypic evolution in birds has also received a lot of attention and in a large compilation of comparative databases, only young clades of birds (clades ≤ 25 million years old) were found to show strong evidence of early bursts of phenotypic evolution (Harmon et al. 2010). Inferring deep time dynamics in lineages that have undergone many recent radiations is a major problem in evolutionary biology, as recent radiations both phenotypically and in terms of lineage diversity can obscure deep time patterns. By combining large-scale analyses of modern birds with estimates of disparity from two major bird Lagerstätten, I am able to document the minimum rate at which birds diversified ecologically, and compare that rate to those estimated via various methods commonly used by evolutionary biologists. This approach provides evidence for the rate of ecological evolution in the early avian radiation, and also provides insight into how different model selection criteria can be used for reliable inferences of deep time dynamics in clades in which fossil data are lacking.

In this article, I estimated the mode of the avian radiation using the “classic” approach of maximizing the likelihood of extant disparity on a time-calibrated molecular phylogeny (Harmon et al. 2010), and using a recently developed Bayesian method that incorporates fossils as node priors (Slater et al. 2012). These methods allowed me to compare the support for the Early Burst

(EB) model to that of Brownian Motion (BM), and also produced parameter estimates. The adequacy and sensitivity of these approaches have recently been called into question (e.g., Boettiger et al. 2012), and so I also leveraged the posterior predictive approach from Slater and Pennell (2013) to again compare models and estimate evolutionary parameters. By measuring a large number of fossil birds from two time periods, I was able to compare the disparity each model-based approach predicted for birds in the past to the observed disparity in the fossil record. Due to biased preservation, the disparity observed in the fossil record is an underestimate of the actual disparity that existed, as many taxa that were alive fail to make it into the fossil record. This means that any comparative method that predicts disparity lower than that observed can be rejected with confidence, whereas methods based on extant taxa that predict variance higher than that observed in the fossil record can be considered consistent to some degree with the independent evidence from the fossil record.

Methods

DATA

I used linear measurements of length of the humerus, ulna, carpometacarpus, femur, tibiotarsus, and tarsometatarsus as well as the widths for all those elements except the carpometacarpus for 604 genera of extant birds (1375 specimens) and 58 genera of extinct birds (202 specimens). Measurements were taken from the right and left sides of every specimen where possible, and averaged between sides for a single specimen, then among sexes within a single species, then among species within a single genus to form a generic average. Generic averages were necessary to ensure that extinct and extant values were comparable (see Mitchell and Makovicky 2014). These linear measurements were combined with categorical beak scores for dentary curvature (straight, ventrally deflected, or fully recurved), culmen curvature (straight, tip recurved, or fully recurved), beak length relative to cranial length (shorter, subequal/equal, longer than cranium, or more than twice cranium length), and finally the width/height ratio of the beak (narrower than tall, as wide as tall, wider than tall). Categorical beak scores were used instead of measurements, as fossil specimens were too crushed to allow for continuous measures. I computed the Gower distance (Gower 1971; Legendre and Legendre 1998) between taxa using these data and an equal weighting scheme for each character, and then used a principal coordinates analysis (PCo) with a Lingoes correction (Lingoes 1971; Legendre and Legendre 1998) to remove negative Eigenvalues. This produced 14 axes, with analyses in this article primarily focused on the first (29% of the variance) and second (17% of the variance) axes (see Inferring Ecology from Morphology section below). This dataset is an expanded version of that published in

Mitchell and Makovicky (2014), and the updated dataset and code for analyzing it is included in the Dryad file associated with this article. Use of single principle coordinate axes without a phylogenetic correction can inflate the signal of models such as EB relative to BM and Orstein–Uhlenbeck artificially (Uyeda et al. 2014), and so the support for EB, especially on axes 1 and 2, should be interpreted as artificially high.

The extinct genera analyzed here are from the Eocene Green River Formation (52 million years ago) and Messel pits (47 million years ago). Both the Green River and Messel have a diversity of small-bodied, perching forms (zygodactylids and primobucconids), as well as numerous aquatic forms (e.g., *Messelornis* and *Limnofregata*). For the datasets here, only described specimens were used for Messel, whereas undescribed but novel forms were included for Green River assemblage if those specimens were housed at an accredited museum (see Dryad file for specimen numbers and measurements). The largest-bodied forms in both Eocene assemblages lack complete enough skeletons for inclusion in this dataset, although large birds are known to be present in the Eocene deposits (e.g., *Vadaravis*, *Masillastega*, *Gastornis*).

INFERRING ECOLOGY FROM MORPHOLOGY

Using morphology as a proxy for ecology is potentially problematic, even for extant taxa. To accurately interpret results, the correlation of morphological factors with ecological factors for extant taxa must be known, and the accuracy of ecological reconstructions in extinct species must be tested against preserved ecological evidence where possible (Zanno and Makovicky 2011; Mitchell and Makovicky 2014). I used binary scores for habitat occupancy and diet in extant birds to test for correlations between morphology and ecology, per Mitchell and Makovicky (2014). The first PCo axis (PCo1: 29% of the variance) is primarily a body size axis, although with some habitat and dietary associations such that negative values represent very small, aerial-foraging, and primarily nectivorous or insectivorous forest-dwelling birds (negative correlations between PCo1 and ecology: -0.31 for forest-dwelling, -0.27 for nectar-eating, -0.35 for arboreal insectivore) whereas large values represent larger-bodied birds (positive correlations between PCo1 and ecology: 0.29 for wetlands, 0.22 for marine, 0.2 for lakes, 0.32 for fish-eating, 0.36 for terrestrial vertebrate-eating, and 0.23 for aquatic arthropod-eating).

To test how well extant ecomorphological relationships predict extrinsic evidence of ecology in extinct forms, I used the method described in Mitchell and Makovicky (2014). This method models the probability an extinct species has a certain ecological trait as a function of the morphological distance between the extinct form and all extant taxa, in which close morphological relationships are taken as evidence of ecological similarity. This can be validated by comparing the probability the extrinsic ecological data associated with an extinct species (e.g., gut contents)

Table 1. The accuracy of PCo1 for predicting ecological traits in Messel birds with known gut contents compared with the probability of predicting those same ecological traits based solely on how common they are in extant birds.

Genus	PCo1 accuracy	Random
<i>Messelornis</i>	0.012	0.108
<i>Primobucco</i>	0.56	0.396
<i>Eocoracias</i>	0.644	0.013
<i>Eoglaucidium</i>	0.639	0.448
<i>Oligocolius</i>	0.56	0.448
<i>Primozygodactylus</i>	0.531	0.448
<i>Colymboides</i>	0.09	0.108
<i>Pumiliornis</i>	0.088	0.053
<i>Selmes</i>	0.585	0.448

Bold indicates the piscivorous taxa *Messelornis* and *Colymboides* are the only genera in which PCo1 underperforms relative to random assignment based on the ecological frequencies in extant birds.

is accurately predicted from the similarities with extant species. Nine described Cenozoic birds from Europe, primarily from the Messel pits, have associated gut contents and are amenable to this test (reviewed in Naish 2014), and the first PCo axis alone was tested for its predictive power (see Table 1).

TAPHONOMIC CONSIDERATIONS

Comparing the morphological variance of fossil assemblages to modern variance is confounded by the imperfect preservation of fossil forms. Specifically, certain ecological traits impart greater preservation potential (e.g., large body size, aquatic habits), and so extinct species with these traits are overrepresented in fossil deposits relative to their true abundances. Mitchell (2015) used a model originally based on habitat filtering (Shipley et al. 2006; Warton et al. 2014) to quantify how ecological traits in birds are related to preservation probability. This model is a Poisson regression that derives the species-specific preservation probability based on the ecological traits and abundance of that species. I used the data from Mitchell (2015) on 53 subfossil assemblages ranging in age from the Late Pleistocene (<100,000 years old) to recent, and expanded on the methods there to predict the expected morphological variance (variance of PCo axis 1) under both a filtering and uniform model of preservation. This facilitates interpretation of deep time deposits by showing how the preserved variance in an ancient assemblage matches what we would expect from a fossilized modern community.

EVOLUTIONARY MODELS

I used 100 trees from the set of trees given in Jetz et al. (2012) that are based on genetic data, pruned to 604 genera with morphological data, to fit three evolutionary models (BM, Ornstein–Uhlenbeck [OU], and EB) via maximum likelihood of the phy-

logenetic variance–covariance (VCV) matrix using the package geiger (Harmon et al. 2008b) for the statistical software R (R Core Team 2014) to explore how phylogenetic uncertainty impacts our ability to understand evolutionary mode. BM is the foundational model of modern comparative methods (Felsenstein 1985) and expects a constant increase in variance through time. OU is a model originally derived from population genetics to detect stabilizing selection (Hansen 1997) but that functionally models a steady state of variance through time (Butler and King 2004). Finally, the EB model is an adaptation of BM, in which the rate of variance increase decays through time at a certain rate (r), and is meant to model adaptive radiations (Harmon et al. 2008, 2010). I also fit these models to a tree with all of the Passeriformes removed, resulting in a tree of 363 extant taxa to explore whether the Neogene radiation of Passeriformes obscures deep time reconstructions of evolutionary patterns in birds. All models were compared using Akaike weights (Burnham and Anderson 2002), as is common in similar analyses. Simply determining which model fits best relative to other models can be highly misleading if none of the models are accurate descriptions of process, or if multiple processes can produce identical patterns. This means that to understand evolutionary dynamics, each model and fit metric should be understood as summarizing different aspects of the pattern of traits along the tree (see below and Discussion).

To understand how incorporating fossil data directly in the model-fitting process influences model support, I included 58 species as node priors (with the nodes determined by the assignment of each species in its most recent assessment; see CenBrackets file in Dryad) using the method described in Slater et al. (2012). This method treats fossil data as a prior on the ancestral state of the closest node, and for most crown clades only a single fossil is known. For those nodes, the morphological value of the extinct species along PCo 1 was used as the mean, and a static SD of 0.1 was used for the prior. For clades with multiple fossils (e.g., the mousebirds; Coliiformes), I created the node prior by using the mean and SD of all taxa close to that crown node. The ages of the nodes were unaltered from the Jetz et al. (2012) trees. Due to the intensive nature of fitting this method, the node prior method was applied to seven of the Jetz et al. (2012) trees.

All of the methods above optimize parameters to fit the VCV matrix of the phylogeny. However, the models can also be assessed by comparing how well the parameters fit other metrics such as the distribution of node heights through time (Freckleton and Harvey 2006) and the partitioning of disparity among and between clades (Harmon et al. 2003). I assessed the fit of the BM and EB models by adjusting the parameters by the fit of various previously developed posterior predictive metrics (Slater and Pennell 2013). These posterior predictive metrics are fit to extant-only trees, and compare the predicted slope between contrast size against node height for a given parameter set to the observed slope found using either

robust regression (this downweights outliers; Robust Linear Model (RLM)) or ordinary least-squares regression (OLS). The slope of the contrast size through time is based on the Nodes Height Test of Freckleton and Harvey (2006). The other metric used to check the fit of parameters is a comparison of the predicted to observed distribution of relative subclade disparity (morphological disparity index, MDI; Harmon et al. 2003). The MDI evaluates what fraction of the total pairwise distance between tips is represented in each subclade. Under an EB model, clades differentiate quickly and converge rarely, resulting in relatively discrete subclades that contain much less of the total disparity than a BM would predict. For these analyses, I measured fit as either the MDI (in which 0 is a perfect fit) or as the log density of the true slope value in the posterior distribution (for RLM and OLS). I computed the harmonic mean of the Akaike Information Criterion corrected (AICc) scores from the Markov-Chain Monte Carlo (MCMC) chain for each of these methods to find which model of morphological evolution was best supported, and then used the parameter estimates for the best-fit models to simulate avian evolution through time. Extant-only trees were used as the fit of the parameters to the observed fossil disparity as an additional form of posterior predictive check.

I also analyzed the fit of model parameters by computing the number of elapsed phylogenetic half-lives for the EB models (Slater and Pennell 2013). The phylogenetic half-life is a way of rendering the rate of rate-decay (r) in the EB model in a more intuitive form, as it describes the number of times the rate of morphological evolution has been halved, computed as $\log(2)/r$ divided by the total depth of the tree.

EVOLUTIONARY SIMULATIONS

Each of the fossil species examined here has been identified as a stem member of a modern clade. For instance, *Tynskya* has been described as a stem psittaciform, and so was placed on the tree at a random (uniform) point along the edge separating the crown psittaciform node (*Nestor* + *Micropsitta*), and the next most ancient node (Psittaciformes + Passeriformes). For crown clades with multiple fossil taxa stem members (e.g., *Selmes* and *Oligocolius* both stem mousebirds), both fossils were placed on the pertinent edge with a node of random length, drawn from a uniform distribution between the maximum age possible given the topology and 0.1 million years ago, separating them (rather than combining them for a distribution, as done above for node priors). A table of fossil taxa and the crown clade they are most closely related to and the code I used for placing them on the phylogeny are included in the Dryad repository.

I used each phylogeny, trimmed to the deposit's age, to simulate trait evolution along the molecular phylogeny. This method produces an ultrametric phylogeny with maximum depth equivalent to the time between the ancestral node in Aves and the age of the fossil deposits. These sliced phylogenies are smaller than

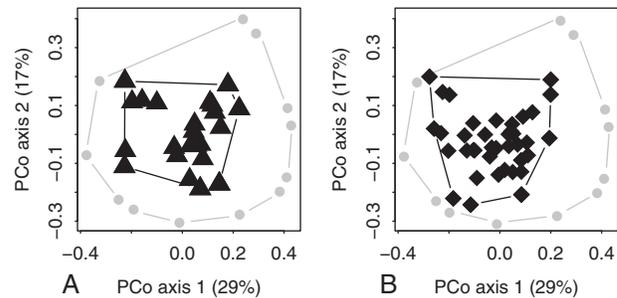


Figure 1. Scatter plots showing the morphological disparity of extant birds (gray outlines) and fossil birds (black points) from the (A) Green River Formation and (B) Messel Pits.

trees produced via birth–death simulations parameterized by the rates inferred from extant birds, but still have far more tips than there are described fossil taxa for each deposit (median number of tips for Green River: 85.5 and Messel: 102), and retain the known topology and internode distances of the molecular tree up to the age of the deposit. The sliced trees were rescaled according to the relevant parameter (r for EB, α for OU, no rescaling for BM) and then I simulated trait evolution along them and computed the variance of the tips. The simulated results for 52 and 47 million years ago were compared to the variance in PCo axis 1 observed in the fossil record to find the fit. Because the fossil deposits represent restricted geographic ranges and the simulations are meant to predict global avian disparity, and because the process of fossilization artificially decreases the variance preserved relative to the variance that was alive (Mitchell 2015), the observed variance at about 50 million years ago was taken as a hard minimum. Any model that failed to predict a level of morphological variance at least as large as the preserved variance was considered inconsistent and rejected, whereas a model that predicted any level of variance at or above the variance observed was considered consistent with the current available data. All datafiles and code necessary to recreate these analyses are included in the Dryad repository associated with this article, as are the model fitting outputs.

Results

The fidelity of ecomorphological relationships in extant taxa was largely upheld when the accuracy of predicting gut contents in extinct species was assessed (Table 1). As found in Mitchell and Makovicky (2014), partially piscivorous species (*Messelornis*, *Colymboides*) were the only group in which extant ecomorphology did not perform better than random assignment (likely due to the categorical nature of the beak data).

The Eocene birds occupy a large proportion of the morphospace defined by extant birds (Fig 1). Ecological biases in

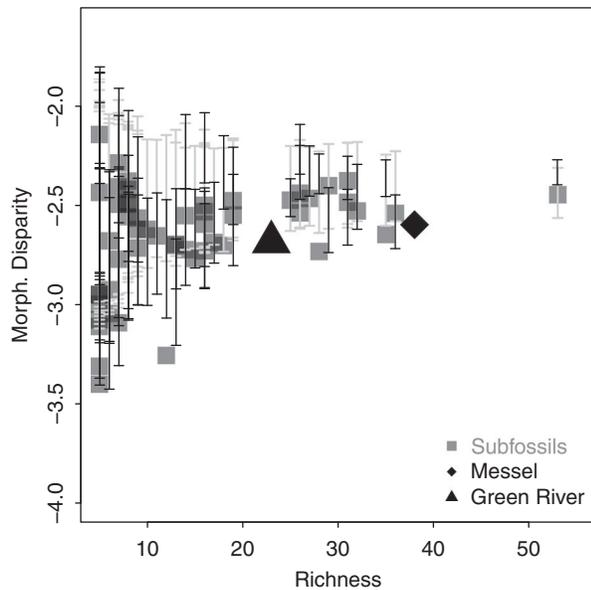


Figure 2. Disparity (log variance in the first morphological PCo axis) for two Eocene fossil Lagerstätten (black points) compared with the Pleistocene-to-recent fossil assemblages (gray squares). Morphological disparity predicted from ecological preservation (dark gray bars) and uniform preservation (light gray bars) is shown to demonstrate the range of morphological disparity for fossilized modern assemblages.

preservation confirm previous reports that, in fossilized bird assemblages, the observed disparity is lower than the true disparity (Fig. 2). This result confirms that the observed variance in morphology for a fossil assemblage is an underestimate of the variance the true, living assemblage had. The Eocene assemblages do show a constricted range, with relatively few small-bodied birds, although large-bodied birds are known from these deposits from scattered remains (see Methods). The disparity in the Eocene Lagerstätten falls well within the bounds of what a modern ecosystem would look like after being preserved in the fossil record.

That the Eocene deposits compare well with taphonomically altered modern assemblages does not mean that the ancient assemblage actually had the same disparity as the modern world, but it does set a minimum bound for the disparity. In other words, the disparity of the bird assemblages that lived in the Eocene was minimally what is preserved, as many taxa that actually lived remain undiscovered or were not preserved at all or adequately enough to be included here (e.g., *Gastornis*), and the subset of taxa that were preserved and have been discovered is a biased view of the total disparity that lived at the time.

To estimate how much disparity is expected in the ancient deposits, I fit the various evolutionary models to the tree of living birds. Akaike weights from the classic approach for the three most common evolutionary models are presented in Figure 3A. For none of the 100 trees from Jetz et al. (2012) is the EB model

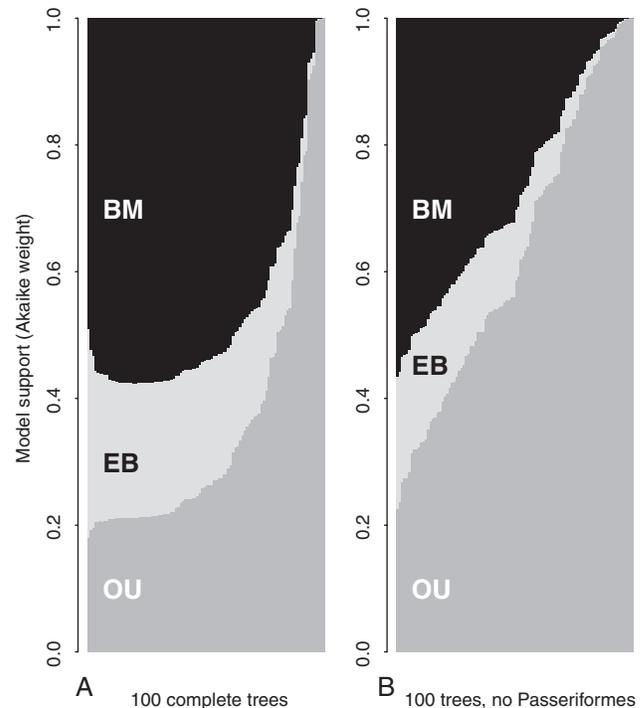


Figure 3. Akaike weights for three different evolutionary models fit over 100 phylogenies, ordered by support for OU, drawn from the genetically constrained trees of Jetz et al. (2012) either including (A) or excluding (B) Passeriformes. Although the relative support of BrM and OU varies substantially, EB is never the best-supported model.

preferred (median Akaike weight: 0.20, max: 0.33), whereas BM has the highest average weight (median: 0.53, max: 0.58) and OU has the highest maximum weight (median: 0.26, max: >0.99). The alpha parameter for the OU model ranges from essentially 0 to 0.027, with the trees in which OU is best supported having the highest alpha values (Fig. 4A). When Passeriformes are excluded, the EB model is supported even less (median weight: 0.11), and the steady-state OU model is the best fit across most trees (median weight: 0.57; Fig. 3B). Support for the EB model was uncorrelated with node age across the different phylogenies (Fig. 5).

Both the likelihoods fit to the VCV matrix and the posterior predictive check using the changes in contrast size through time are more consistent with a BM model. Of all the extant-only fit metrics used, only the comparison of the observed subclade disparity to that predicted from the various parameter configurations supports the EB model (median rate-decay parameter, r : -0.0019; Fig. 6). The methods based only on extant taxa (classic approach and posterior predictive method) suggest that almost no phenotypic half-lives have elapsed. This means that the best-fit EB pattern is still nearly indistinguishable from a general BM model (Fig. 6). However, weakly incorporating fossil data using the method of Slater et al. (2012) supports a much larger number

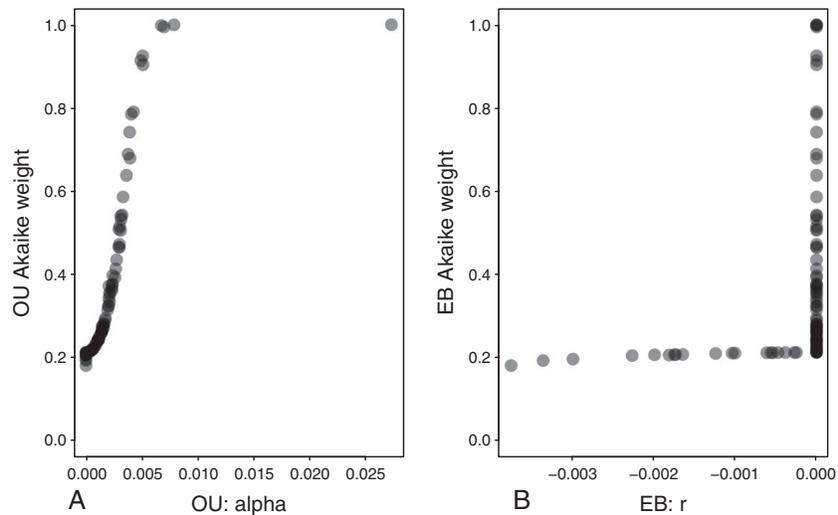


Figure 4. Support (Akaike weight) for the OU model (A) and EB model (B) plotted against their non-Brownian parameters (alpha and r).

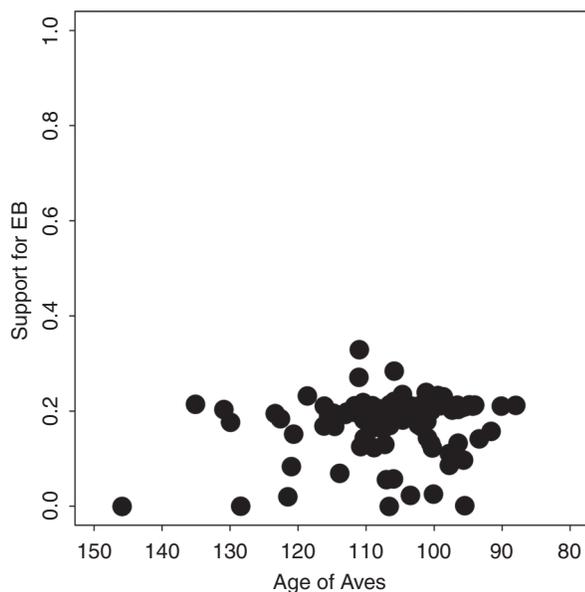


Figure 5. Support (Akaike weight) for the EB model plotted against the age of crown Aves in the 100 different phylogenies examined here. Support for EB does not vary with node age, consistent high recent rates of evolution being largely responsible for the poor fit of EB.

of half-lives elapsing (due to a higher rate of phenotypic decay; median r parameter: -0.049 ; Fig. 7), supporting a more dramatic early radiation of birds than the extant-only models. For six of the seven trees the node prior method was applied to, EB was the favored model with an Akaike weight >0.99 . In general, incorporating fossils as node priors produces support for a sufficient amount of morphological evolution early in bird history to account for the high disparity observed at the “halfway point” of avian evolution.

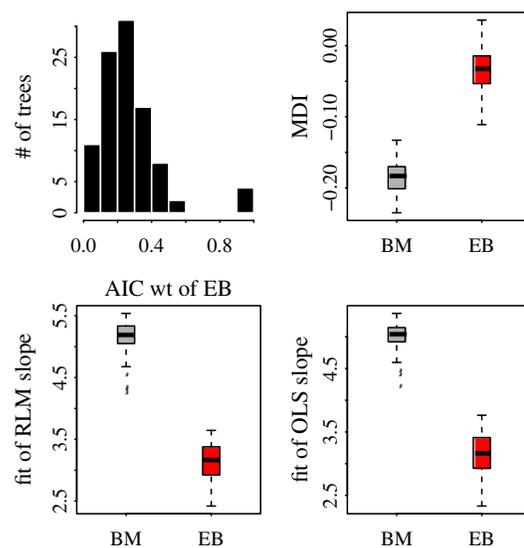


Figure 6. Four metrics of “fit” from Slater and Pennell (2013). Akaike weights are computed using the AICm based on the harmonic mean of the post-burnin (75,000 generations) likelihoods output by the MCMC chain (G. Slater, pers. comm.). For the subclade disparity (MDI), a perfect fit is 0 difference between modeled and observed partitioning, and the figure shows the log density of the posterior distribution at 0 for each tree. For the OLS and robust-regression analyses, the figure shows the log density of the posterior distribution at the observed slope for each tree. Only the subclade disparity partitioning metric shows evidence for the early burst during the first phases of avian evolution observed in the fossil record.

The range of morphologies seen in the fossil assemblages is slightly higher than what is expected from the extant-only models, even when accounting for uncertainty in ancestral state reconstructions (phenogram of the tree with the highest support for EB, and thus the highest ancient variance of all 100 Jetz et al. trees,

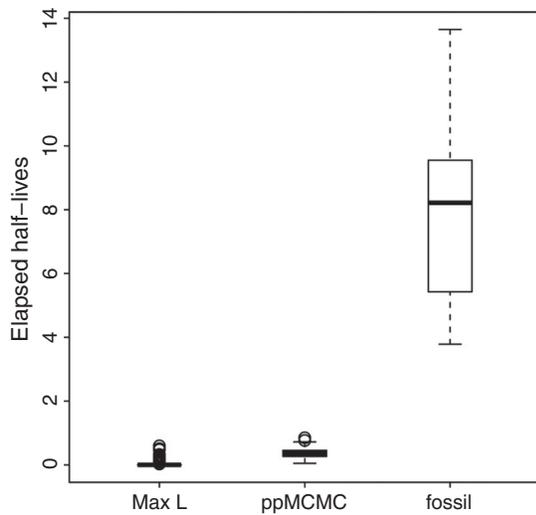


Figure 7. The number of elapsed half-lives (age of the tree's deepest node divided by $\log 2/r$, where r is the rate of rate-decay in the EB model) estimated via maximum likelihood on the extant-only tree from the classic approach (Max L), the posterior predictive approach (ppMCMC), and the fossil informed approach (fossil) are shown. Fewer elapsed half-lives suggest a slower pattern of rate change through time, with zero elapsed half-lives indistinguishable from a BM model.

shown in Fig. 8). Parameters estimated from the extant-only models fail to predict the variance in morphology observed at 52 or 47 million years ago (Fig. 9). The method of using fossil data as node priors produces estimates of ancient variance substantially higher than observed in the fossil record, yet the same parameters produce estimates of extant variance several orders of magnitude higher than found in the extant data (mean variance estimate is 20 times higher).

Discussion

The Green River and Messel bird assemblages are composed entirely of crown avians, and preserve a large proportion of the observed morphological disparity in extant birds generally. The disparity of the Eocene assemblages is indistinguishable from the disparity of simulated fossilized modern assemblages and observed recent subfossil deposits (Pleistocene–Holocene) of comparable species richness. Further, both the Eocene assemblages preserve an array of water birds, ranging from long-legged waders (e.g., *Messelornis*, *Presbyornis*, *Rhynchaetites*) to soaring taxa (e.g., *Limnofregata*, *Masillastega*). When considering only the taxa with complete skeletons, as done here, the disparity of these Eocene deposits is comparable to fossilized modern ecosystems, complete even with nectivorous taxa such as *Pumiliornis*. Large-bodied taxa are absent due to incomplete skeletal preservation, and small-bodied taxa are present at reduced richness, possibly due to taphonomic factors.

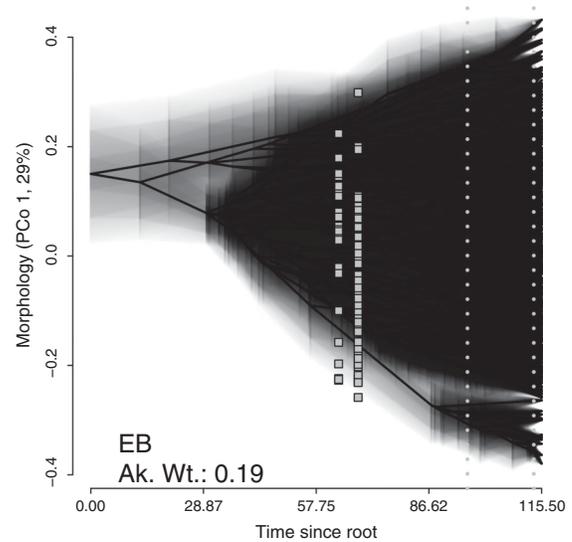


Figure 8. Phenogram of the evolutionary tree from Jetz et al. (2012) with the highest support for the EB model showing the evolution of PCo axis 1 against the time since root. Confidence intervals for the variance represented by semitransparent lines, and slightly underestimate the variance of small-bodied birds, even for this phenogram with the highest ancient variance of all 100. Solid symbols represent observed fossil specimens, and the dashed lines represent the time period for which Harmon et al. (2010) found strong support for the EB model.

The difference between the Eocene and modern assemblages is mostly due to the lower diversity of small-bodied terrestrial forms in the Eocene, which are represented by only a handful of taxa (*Primobucco*, zygodactylids, and stem coliid). Unfortunately, a large proportion of specimens from both Lagerstätten are held in private collections, making a comprehensive assessment of abundance impossible, but it is still clear from the specimens housed in public institutions that the zygodactylids, stem coliid (mousebirds), and *Primobucco* are extremely abundant despite their small size. The low richness but high abundance of these small forms is consistent with taphonomic expectations, as the only small-bodied birds that are likely to be preserved are those with ecological traits that mitigate the bias against their preservation (e.g., passerines that nest near water, like many blackbirds, are preserved in high abundance in the fossil record; Mitchell 2015).

Excluding Passeriformes from the extant tree produced much higher support for the OU model of morphological evolution. This suggests that birds reached the bounds of their total morphospace quickly, and have been exploring it at a more-or-less consistent rate ever since. This model fitting nonpasserine birds is actually consistent with the verbal description of an EB, as it suggests morphospace was saturated quickly, but not necessarily in the style of a classic adaptive radiation, as there is no evidence for a slowdown in rates. However, the discreteness of different

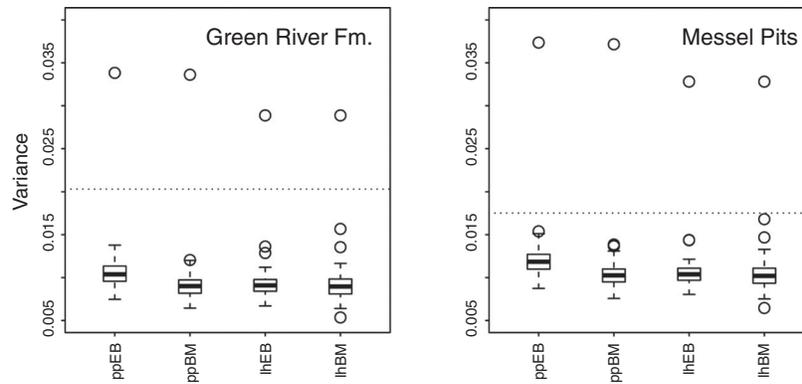


Figure 9. Variance along the first morphological PCo axis for 100 simulations using the parameters from the posterior predictive and maximum-likelihood approach for the Green River (A, left) and Messel (B, right). Horizontal lines represent the disparity observed in each Lagerstätten. For all simulations, there was one tree of the 100 that produced extremely high estimates of variance, and that is the tree figured as a phenogram in Figure 7.

avian clades (evidenced by the MDI) is inconsistent with the OU model. As, if birds were exploring morphospace freely through time, there would be no pronounced differences between ancient clades (e.g., ratites and anseriforms would be unlikely to be as morphologically differentiated as they are). This raises an important point about the application of comparative models, as most models were built with an inherent mechanistic basis (e.g., OU as adaptation and selection, EB as adaptive radiation); however these models only quantify patterns that could have been produced by multiple mechanisms. For instance, OU will fit better than most other models with recently diverged lineages are very divergent relative to the total variance of the clade. This could happen due to stabilizing selection, as proposed, but also due to divergent selection between close relatives resulting from ecological character displacement or due to functional or developmental constraints that limit the total variance achievable by that clade, for example.

The variance and range in morphology observed in the fossil assemblages from the “halfway point” of avian evolution is ~70% of the modern, which is substantially higher than models based solely on extant taxa would predict for ~50Ma. This observation, of crown Aves having achieved such ecological disparity by the Eocene, stands in stark contrast to expectations from modern data alone. None of the models based on extant taxa only consistently predicted this high level of early disparity, and the fossil-informed method was unable to predict both the high level of ancient disparity and the relatively low modern disparity simultaneously. All of these models are known as extreme simplifications, but they are commonly used to at least predict the broad contours of morphological evolution.

The “classic” approach of fitting rates of phenotypic evolution directly to the VCV matrix had low support for the EB model, and estimated parameters of rate change so low that almost no half-lives went by. Using fossils as node priors supports a rate change, but it results in estimated rates of phenotypic evo-

lution too high to predict the low level of disparity observed in the modern world. The high variance in the Eocene deposits contrasts sharply with the low support for higher rates of morphological evolution found using common comparative methods. These results above support simulations that suggest that most comparative methods have low power to detect deep time EB patterns (Boettiger et al. 2012). However, the MDI (Harmon et al. 2003) was one metric capable of inferring high, early rates of morphological evolution from extant taxa only. For evolutionary biologists interested in exploring dynamics in clades without well-resolved fossil records, the partitioning of disparity among subclades seems the most robust in terms of these simple models, even if the estimated parameters cannot recapitulate the patterns of the fossil data.

A major reason the parameter estimates for the EB model fit to large clades are likely to be inaccurate is due to extremely rapid recent radiations, as small subclades undergo dramatic radiations (e.g., Darwin’s finches). Rapid divergences among closely related species are highly unlikely at high levels of rate decay, despite the expectation that iterative radiations should occur as ecological conditions change, as when subclades invade new biogeographic provinces, to produce a series of multiple radiations (see Hopkins and Smith 2015, for an example of hierarchical radiations). More complex future models that are informed by ecology may circumvent this, by, for instance, resetting the rates of decay each time a clade invades a new biogeographic province or niche space. Modern comparative methods are based on the insights of Felsenstein (1985), in which species are not evolutionarily independent until they diverge. The observation that lineages are perfectly nonindependent before divergence is true and necessary; however, the assumption that once lineages diverge, their evolution is completely independent, essentially ignores all ecological principles, although comparative biologists are working to correct this (e.g., Ingram et al. 2012).

Phylogenetic uncertainty can also have a large impact on estimates of model fit, as even if the topology is correct, if nodes are dated in a systematically biased way (e.g., if the ages of younger nodes are more accurate than older nodes), then the model fitting can be positively misled. By using the more reliable, genetic-based trees of Jetz et al. (2012) and examining fit over a suite of trees with differing topologies and node ages, the difficulty of detecting an early burst does not appear to be an artifact of uncertainty alone. However, the ages of the nodes along the Jetz et al. (2012) tree now appear to be systematically too old, based on recent genomic results (Jarvis et al. 2014).

Both the likelihoods based on the VCV matrix of the phylogeny and the posterior predictive check of the regression of contrast size against time suggest that recent clades have much higher rates of phenotypic evolution than would be expected if a strict EB model held through the history of Aves. However, the disparity is partitioned among subclades in a way that supports an early burst pattern, as different subclades converge much less than would be expected if all taxa were evolving independently (i.e., individual subclades represent a smaller fraction of the total disparity than expected under BM). As all of our models of phenotypic evolution are highly simplified, it is better to think about what aspect of evolution each fit metric describes, rather than try to interpret them as evidence for or against a single specific evolutionary mechanism. When these different metrics are combined with the high disparity observed in the fossil record, a highly heterogeneous picture of avian evolution can be inferred.

Early in their history, birds underwent rapid ecological evolution, and different subclades became distinct and have converged relatively rarely (i.e., less than expected under a BM or OU model). As birds expanded and invaded new regions, and species went extinct, younger adaptive radiations have produced extremely high rates of recent morphological evolution (i.e., higher than a strong EB model would predict). These subsequent radiations show high enough rates of evolution to reject a model in which rates decline monotonically through time; however the early radiation was profound enough to both generate high ancient disparity (as seen in the fossil deposits), and to leave a signal in how subclades partition the overall disparity (as seen in the MDI).

The fossil record was long the only way to reconstruct ancient patterns of morphological evolution. With the advent of modern comparative methods (Felsenstein 1985), evolutionary biologists began attempting to infer deep time dynamics from extant records only. Simulations and empirical results have shown that comparative methods have both low power (Boettiger et al. 2012) and an inability to predict nonmonotonic changes in disparity. As more complex ecological models come to the forefront, we may be better able to capture the more realistic, nonmonotonic dynamics of morphological evolution (Slater 2013). Considering how much

of life's modern diversity is sequestered in clades without good fossil records, this seems to shine a pessimistic light on our collective endeavor to understand the history of life. However, by using clades with fossil data to find and fit reliable methods for inferring morphological evolution (e.g., subclade disparity through time), paleontologists and comparative biologists can work together to bring our data to as close to representative as possible.

ACKNOWLEDGMENTS

This article was greatly improved by comments from M. Friedman, M. Hopkins, an anonymous reviewer, G. Slater, D. Bapst, P. Makovicky, M. Foote, K. Angielczyk, M. Webster, D. Rabosky, and T. Sosa. Discussion with members of the Field Museum's Bird Division and use of the collections at the Field Museum, American Museum, Wyoming Dinosaur Center, Smithsonian Institution, Utah Museum of Natural History, Royal Museum of Belgium, Senckenberg Museum, and the Berlin Museum also improved this article. This work was financially supported by the Hinds Fund, and NSF grants EAPSI 1107676 and DDIG 1311389.

DATA ARCHIVING

The doi for my data is 10.5061/dryad.p2d8k.

LITERATURE CITED

- Barker, F. K. 2011. Phylogeny and diversification of modern passerines, in *Living Dinosaurs: The Evolutionary History of Modern Birds* (eds G. Dyke and G. Kaiser). John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9781119990475.ch9
- Behrensmeyer, A. K., S. M. Kidwell, and R. A. Gastaldo. 2000. Taphonomy and paleobiology. *Paleobiology* 26:103–147.
- Behrensmeyer, A. K., C. T. Stayton, and R. E. Chapman. 2003. Taphonomy and ecology of modern avifaunal remains from Amboseli Park, Kenya. *Paleobiology* 29:52–70.
- Benson, R. B., and J. N. Choiniere. 2013. Rates of dinosaur limb evolution provide evidence for exceptional radiation in Mesozoic birds. *Proc. R. Soc. Lond. B Biol. Sci.* 280:20131780. Available at <http://dx.doi.org/10.1098/rspb.2013.1780>.
- Boettiger, C., G. Coop, and P. Ralph. 2012. Is your phylogeny informative? Measuring the power of comparative methods. *Evolution* 66:2240–2251.
- Brown, J., J. Rest, J. Garcia-Moreno, M. Sorenson, and D. Mindell. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biol.* 6:6.
- Brusatte, S. L., G. T. Lloyd, S. C. Wang, and M. A. Norell. 2014. Gradual assembly of avian body plan culminated in rapid rates of evolution across the dinosaur-bird transition. *Curr. Biol.* 24:2386–2392.
- Burnham, K. P., and D. R. Anderson. 2004. *Model selection and multimodel inference: a practical information theoretic approach*, 2nd edn. Springer, New York.
- Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164:683–695.
- Feduccia, A. 1995. Explosive evolution in tertiary birds and mammals. *Science* 267:637–638.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- Finarelli, J. A., and J. J. Flynn. 2006. Ancestral state reconstruction of body size in the Caniformia (Carnivora, Mammalia): The effects of incorporating data from the fossil record. *Syst. Biol.* 55:301–313.
- Foote, M. 1997. The evolution of morphological diversity. *Ann. Rev. Ecol. Evol.* 28:129–152.

- Freckleton, R. P., and P. H. Harvey. 2006. Detecting non-Brownian trait evolution in adaptive radiations. *PLoS Biol* 4:e373.
- Gower, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27:857–871.
- Grande, L. 2013. *The lost world of Fossil Lake: snapshots from deep time*. Univ. of Chicago Press, Chicago, Illinois.
- Hackett, S. J., R. T. Kimball, S. Reddy, R. C. K. Bowie, E. L. Braun, M. J. Braun, J. L. Chojnowski, W. A. Cox, K.-L. Han, J. Harshman et al. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320:1763–1768.
- Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Am. Nat.* 51:1341–1353.
- Harmon, L. J., J. A. Schulte, A. Larson, and J. B. Losos. 2003. Tempo and mode of evolutionary radiation in Iguanian lizards. *Science* 301:961–964.
- Harmon, L. J., J. Melville, A. Larson, and J. B. Losos. 2008a. The role of geography and ecological opportunity in the diversification of day geckos (*Phelsuma*). *Syst. Biol.* 57:562–573.
- Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008b. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24:129–131.
- Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. Bryan Jennings, K. H. Kozak, M. A. McPeck, F. Moreno-Roark, T. J. Near et al. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385–2396.
- Hopkins, M. J., and A. B. Smith. 2015. Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution. *Proc. Natl. Acad. Sci. USA* 112:3758–3763.
- Ingram, T., L. J. Harmon, and J. B. Shurin. 2012. When should we expect early bursts of trait evolution in comparative data? Predictions from an evolutionary food web model. *J. Evol. Biol.* 25:1902–1910.
- James, H. F. 2005. Paleogene fossils and the radiation of modern birds. *Auk* 122:1049–1054.
- Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. W. Ho, B. C. Faircloth, B. Nabholz, J. T. Howard et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.
- Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. *Nature* 491:444–448.
- Ksepka, D. T., and C. A. Boyd. 2012. Quantifying historical trends in the completeness of the fossil record and the contributing factors: an example using Aves. *Paleobiology* 38:112–125.
- Ksepka, D. T., J. A. Clarke, S. J. Nesbitt, F. B. Kulp, and L. Grande. 2013. Fossil evidence of wing shape in a stem relative of swifts and hummingbirds (aves, pan-apodiformes). *Proc. R. Soc. Lond. B Biol. Sci.* 280:20130580. <http://dx.doi.org/10.1098/rspb.2013.0580>
- Legendre, P., and L. Legendre. 1998. *Numerical ecology*. 2nd English ed. Elsevier, Amsterdam.
- Lingoes, J. C. 1971. Some boundary conditions for a monotone analysis of symmetric matrices. *Psychometrika* 36:195–203.
- Livezey, B. C., and R. L. Zusi. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. ii. Analysis and discussion. *Zool. J. Linn. Soc.* 149:1–95.
- Manegold, A. 2008. Passerine diversity in the late Oligocene of Germany: earliest evidence for the sympatric coexistence of Suboscines and Oscines. *Ibis* 150:377–387.
- Mayr, G. 2005. The paleogene fossil record of birds in Europe. *Biol Rev.* 80:515–542.
- Mayr, G., and A. Manegold. 2004. The oldest European fossil songbird from the Early Oligocene of Germany. *Naturwissenschaften* 91:173–177.
- Mayr, G., and V. Wilde. 2015. Eocene fossil is earliest evidence of flower-visiting by birds. *Biol. Lett.* 10:20140223
- Mitchell, J. S. 2015. Preservation is predictable: quantifying the effect of taphonomic biases on ecological disparity in birds. *Paleobiology* 41:353–367.
- Mitchell, J. S., and P. J. Makovicky. 2014. Low ecological disparity in Early Cretaceous birds. *Proc. R. Soc. B Biol. Sci.* 281:20140608. <http://dx.doi.org/10.1098/rspb.2014.0608>.
- Morlon, H., T. L. Parsons, and J. B. Plotkin. 2011. Reconciling molecular phylogenies with the fossil record. *Proc. Natl. Acad. Sci. USA* 108:16327–16332.
- Martin Hughes, Sylvain Gerber, and Matthew Albion Wills. 2013. Clades reach highest morphological disparity early in their evolution *PNAS* 110(34):13875–13879. doi:10.1073/pnas.1302642110
- Naish, D. 2014. The fossil record of bird behaviour. *J. Zool.* 292:268–280.
- Nesbitt, S. J., D. T. Ksepka, and J. A. Clarke. 2011. Podargiform affinities of the enigmatic *Fluvioviridavis platyrhamphus* and the early diversification of Strisores (“Caprimulgiformes” + Apodiformes). *PLoS ONE* 6:e26350.
- Pacheco, M. A., F. U. Battistuzzi, M. Lentino, R. F. Aguilar, S. Kumar, and A. A. Escalante. 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. *Mol. Biol. Evol.* 28:1927–1942.
- Pant, S. R., A. Goswami, and J. A. Finarelli. 2014. Complex body size trends in the evolution of sloths (*Xenarthra: Pilosa*). *BMC Evol. Biol.* 14. Available at dx.doi.org/10.1186/s12862-014-0184-1.
- R Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Schweizer, M., S. T. Hertwig, and O. Seehausen. 2014. Diversity versus disparity and the role of ecological opportunity in a continental bird radiation. *J. Biogeogr.* 41:1301–1312.
- Shipley, B., D. Vile, and E. Garnier. 2006. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science* 314:812–814.
- Slater, G. J. 2013. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods Ecol. Evol.* 4(8):734–744.
- Slater, G. J., and M. W. Pennell. 2013. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. *Syst. Biol.* 63:293–308.
- Slater, G. J., L. J. Harmon, and M. E. Alfaro. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution* 66:3931–3944.
- Smith, N. D., L. Grande, and J. A. Clarke. 2013. A new species of Threskiornithidae-like bird (Aves, Ciconiiformes) from the Green River Formation (Eocene) of Wyoming. *J. Vertebr. Paleontol.* 33:363–381.
- Turvey, S. T., and T. M. Blackburn. 2011. Determinants of species abundance in the Quaternary vertebrate fossil record. *Paleobiology* 37:537–546.
- Uyeda, J. C., D. S. Caetano, and M. W. Pennell. 2014. Comparative analysis of principal components can be misleading. Available at <http://dx.doi.org/10.1101/007369>.
- Warton, D. I., B. Shipley, and T. Hastie. 2014. CATS regression – a model-based approach to studying trait-based community assembly. *Methods Ecol. Evol.* 6:389–398.
- Zanno, L. E., and P. J. Makovicky. 2011. Herbivorous ecomorphology and specialization patterns in theropod dinosaur evolution. *Proc. Natl. Acad. Sci. USA* 108:232–237.

Associate Editor: M. Friedman
 Handling Editor: M. Servedio