Paleobiology, 37(1), 2011, pp. 1–22

Phylogenetic corrections for morphological disparity analysis: new methodology and case studies

Stephen L. Brusatte, Shaena Montanari, Hong-yu Yi, and Mark A. Norell

Abstract.—Taxonomic diversity and morphological disparity are different measures of biodiversity that together can describe large-scale evolutionary patterns. Diversity measures are often corrected by extending lineages back in time or adding additional taxa necessitated by a phylogeny, but disparity analyses focus on observed taxa only. This is problematic because some morphologies required by phylogeny are not included, some of which may help fill poorly sampled time bins. Moreover the taxic nature of disparity analyses makes it difficult to compare disparity measures with phylogenetically corrected diversity or morphological evolutionary rate curves. We present a general method for using phylogeny to correct measures of disparity, by including reconstructed ancestors in the disparity analysis. We apply this method to discrete character data sets focusing on Triassic archosaurs, Cenozoic carnivoramorph mammals, and Cretaceous–Cenozoic euarchontogliran mammals. Phylogenetic corrections do not simply mirror the taxic disparity patterns, but affect the three analyses in heterogeneous ways. Adding reconstructed ancestors can inflate morphospace, and the amount and direction of expansion differs depending on the taxonomic group in question. In some cases phylogenetic corrections give a temporal disparity curve indistinguishable from the taxic trend, but in other cases disparity is elevated in earlier time intervals relative to later bins, due to the extension of unsampled morphologies further back in time. The phylogenetic disparity curve for archosaurs differs little from the taxic curve, supporting a previously documented pattern of decoupled disparity and rates of morphological change in dinosaurs and their early contemporaries. Although phylogenetic corrections should not be used blindly, they are helpful when studying clades with major unsampled gaps in their fossil records.

Stephen L. Brusatte, Hong-yu Yi, and Mark A. Norell. Division of Paleontology, American Museum of Natural History, New York, New York 10024, and Department of Earth and Environmental Sciences, Columbia University, New York, New York. E-mail: sbrusatte@amnh.org, hyi@amnh.org, norell@amnh.org
Shaena Montanari. Division of Paleontology, American Museum of Natural History, New York, New York 10024. E-mail: smontanari@amnh.org

Accepted: 29 June 2010

Introduction

Taxonomic diversity and morphological disparity are distinct measures of biodiversity, which together can describe large-scale patterns of macroevolution and the tempo of evolutionary radiations over deep time (e.g., Foote 1993, 1996, 1997; Erwin 2007). Taxonomic diversity has traditionally been assessed by counting the number of taxa observed in the fossil record during discrete intervals of time (e.g., Raup 1972; Sepkoski 1984; Alroy et al. 2008), the so-called taxic approach. However, it has long been recognized that the fossil record is biased and that knowledge of a clade's phylogeny can play a correcting role in diversity estimates (e.g., Fisher 1982; Paul 1982). In this ''phylogenetic'' approach, lineage extensions and entirely new lineages posited by a phylogeny, but unsampled in the fossil record, can be added to the observed measurements, producing a phylogenetically corrected diversity estimate (Norell 1992, 1993; Norell and Novacek 1992; Smith 1994). This method is now commonplace in many studies of taxonomic diversity, especially those focusing on extinct vertebrate taxa, which have been the subject of more intense morphological phylogenetic studies than most invertebrate groups (e.g., Lloyd et al. 2008; Barrett et al. 2009; Young et al. 2010).

However, morphological disparity—the variability in shape, form, and body plan among organisms—is universally treated as a ''taxic'' concept: disparity analyses measure the variability of observed taxa with directly observable temporal ranges only (e.g., Wills et al. 1994; Ciampaglio et al. 2001). In large part, this is due to the difficulty of correcting for morphology, as opposed to lineages. It is

0094-8373/11/3701–0001/$1.00

straightforward to extend a lineage further back in time or count an unsampled lineage implied by a phylogeny, but how should morphology be extended or entirely new morphologies be introduced? Despite these difficulties, it is an unavoidable reality that phylogenies often require certain unsampled lineages, which would have their own morphologies, to have existed at certain times. In short, unsampled morphologies are a logical consequence of some phylogenies. If done correctly, including these ''ghost morphologies'' in a disparity analysis should result in more complete, reasonable, and accurate measures of morphological variability.

Furthermore, in practical situations, the taxic nature of disparity estimates may be problematic for three main reasons, and including phylogenetic corrections may help remedy these problems. First, taxic estimates may be prone to major biases in the fossil record, especially entire intervals that are poorly sampled (Smith 1994; but see Sepkoski and Kendrick 1993; Wagner 2000a,b; Lane et al. 2005). In these cases, disparity may appear low simply because only a small number of organisms, and thus a limited amount of morphological variability, are observed. Alternatively, variance-based disparity may appear artificially high if only a few very different morphologies are sampled. In such cases, phylogenetic corrections may help fill gaps in the observed record. Second, the taxic nature of disparity analyses makes it less straightforward to compare disparity curves with phylogenetically corrected diversity curves, a potentially critical problem because the comparison of diversity and disparity over time is at the root of many macroevolutionary studies (e.g., Foote 1993; Wesley-Hunt 2005; Brusatte et al. 2008b; Young et al. 2010).

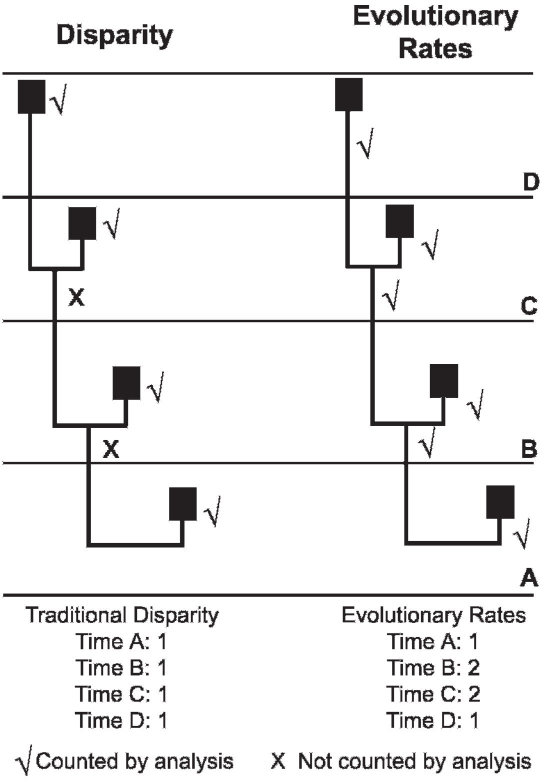

FIGURE 1. A schematic comparison of how analyses of traditional disparity (a ''taxic'' measure) and morphological evolutionary rates (a ''phylogenetic'' measure) bin taxa into temporal bins. Because evolutionary rates analyses (sensu Wagner 1997; Ruta et al. 2006; Brusatte et al. 2008a) consider each branch in the phylogeny, including both external branches leading to observed taxa and internal branches linking hypothetical ancestors, they will always tally more ''counts'' than traditional disparity analyses, which only consider observed taxa.

Third, the taxic nature of disparity measurements also makes it difficult to compare disparity with rates of morphological character evolution (Fig. 1). One common rate metric uses a phylogeny to measure the rate of change on each branch of the tree (number of characters changing divided by time duration of the branch), and branches are then binned by time and averaged to construct a temporal rate curve (Wagner 1997; Ruta et al. 2006; Brusatte et al. 2008a). Because this procedure measures the rate of each branch of the tree, including internal branches that connect hypothetical ancestors that are not observed in the fossil record, it is a phylogenetic approach. Comparison of disparity and rates (as defined above) has only been done occasionally in the literature (Wagner 1997; Brusatte et al. 2008a), but Brusatte et al. (2008) argued that such a comparison is an important and insightful method for characterizing major evolutionary radiations. In fact, they reported an interesting result when they compared disparity and rates in Triassic archosaurian reptiles, including the oldest and most primitive dinosaurs and their early contemporaries. They noted that rates and disparity are decoupled: morphological rates were high early in archosaur history and decreased throughout the Late Triassic, whereas disparity gradually rose during this time. If true, this pattern suggests that rapid rates of character change do not necessarily translate into a wide range of body plans (and potentially diets and lifestyles if morphology is an indicator of ecology), perhaps an unexpected result. However, it is possible that this discrepancy is a figment of improper comparison between taxic disparity and phylogenetic rate measures.

Recognizing both the logical reality of unobserved morphologies and the potential practical issues with taxic measures of morphological disparity, we present a general method for using phylogeny to correct measures of disparity. This approach uses a phylogeny to reconstruct the character states of each hypothetical ancestor (internal node) and then treats these ancestors as real taxa with real character scores, which then factor into the calculations of morphological variability. In effect, we assume that ancestors are real, but cannot be sampled as discrete taxa in the fossil record, and their morphology can be inferred by reference to their descendants. We reanalyze the data set of Brusatte et al. (2008a) with these methods, which should help determine whether an improper comparison between taxic and phylogenetic approaches led to the noted discrepancy between disparity and rates. We also analyze two other data sets, both focusing on fossil mammal clades with well understood phylogenetic relationships, and note how phylogenetic corrections alter the traditional, taxic disparity measures, including both temporal disparity curves and disparity comparisons between taxonomic groups. We then argue that phylogenetic corrections may, in some cases, help alleviate sampling biases and may be an important tool for understanding the morphological variability present during the early history of major clades.

Methods

Archosauria: Reanalyzing the Data Set of Brusatte et al. (2008a)

The primary aim of this study is to provide a general, reasonable method of phylogenetic corrections to disparity analyses. These can then be used to assess whether the discordance between disparity and evolutionary rates found by Brusatte et al. (2008a) in Triassic archosaurs is due to an improper comparison between phylogenetic and taxic methods. Therefore, in all cases we have aimed to retain the methodological protocols outlined by Brusatte et al. (2008a). We utilize the same data set, the same phylogenetic hypothesis, and the same analytical and statistical methodology used in the original study. As details of the original methodology are mostly restricted to an online supplementary appendix that may not be easily available (www.sciencemag.org/cgi/content/full/321/5895/1485/DC1), we summarize the methods here for clarity.

Taxonomic Scope and Aims.—Brusatte et al. (2008a) focused on the morphological evolution of Triassic archosaurs. Archosauria is a speciose and morphologically diverse clade of reptiles that includes living crocodiles and birds, as well as traditional dinosaurs and many subgroups (phytosaurs, aetosaurs, ornithosuchids, ''rauisuchians'') restricted to the Mesozoic (Benton 2005). The two primary goals of the study were to (1) quantify temporal patterns in archosaur evolution, and (2) compare the morphological evolution of dinosaurs and their early contemporaries, in order to assess long-standing ideas of competitive superiority as an explanation for dinosaur ascendancy during the Triassic (see review in Langer et al. 2010 and Brusatte et al. 2010b). Brusatte et al. (2008a) compiled a taxon set of 61 Triassic archosaurs and three close Triassic outgroups. The ingroup taxa included several representatives of each major archosaur subgroup, which were chosen on the basis of completeness, morphological variety (i.e., an attempt to include representatives of each major body plan), and accessibility for first-hand observation. Genera were used as proxies for species, as nearly all Triassic archosaur genera are monospecific. The genera include 24 members of the ''bird-line'' of archosaur phylogeny (1 basal ornithodiran, 4 pterosaurs, 8 non-dinosaurian dinosauromorphs, 11 dinosaurs) and 37 members of the ''crocodile-line'' (4 phytosaurs, 4 aetosaurs, 2 ornithosuchids, 4 crocodylomorphs, 20 ''rauisuchians,'' 3 Crurotarsi *incertae sedis*).

Quantifying Morphology.—The 64 taxa were scored for 437 discrete characters, which cover all elements of the skeleton and together comprise a nearly complete spread of features that evolved within the group during the Triassic. The characters were assimilated from a number of cladistic data sets, including a novel analysis of higher-level archosaur phylogeny (Brusatte et al. 2010a). Ten characters were ordered and the remainder unordered. Autapomorphies of individual genera were not included, because genera are exemplars and including such characters may bias the data set if certain exemplars have more or fewer autapomorphies than average for their subgroup. Additionally, as reconstructed ancestors cannot have autapomorphies by definition, but the real ancestral morphologies they represent would likely have possessed unique characters, it is not evenhanded to include autapomorphy-less ancestors and terminal taxa with autapomorphies in the same data set. However, the data set does include autapomorphies for terminal branches, because it includes characters that diagnose larger groups for which there is only one exemplar. Exemplar usage is consistent across the tree, and major clades and time periods are represented consistently, so this should not introduce a strong bias.

Phylogenetic Framework.—Brusatte et al. (2008a) used a single, resolved cladogram as the basis for their evolutionary rates analysis. Because a primary goal of the current study is to compare disparity and rates, we utilize the same single phylogenetic tree, which is figured in the supplementary information of Brusatte et al. (2008a). This tree was the first most parsimonious tree (i.e., shortest tree) recovered by a phylogenetic analysis of higher-level archosaur relationships (Brusatte 2007). Ideally, several trees would be used to calculate evolutionary rates and/or phylogenetic disparity corrections, in order to better account for topological differences and possible error. However, disparity analysis proceeds in several steps, each of which requires use of different software packages, and there is as yet no automated protocol for combining these analyses. Thus, it is remarkably time consuming to analyze multiple trees, especially as there are multiple character optimization procedures and disparity calculations for each tree used. However, we have experimented with alternative trees, including a new most parsimonious reconstruction of higher-level archosaur phylogeny reported by Brusatte et al. (2010) that differs slightly from that used in the original study, and these give indistinguishable evolutionary rates results.

Reconstructed Ancestors.—Our phylogenetic correction necessitates the introduction of reconstructed ancestors (internal nodes of the phylogeny) into the disparity data set. This is analogous to the inclusion of internal branches in the evolutionary rates data set. In order to include such ancestors, we must determine their character state scores for each of the 437 morphological characters. Ancestral states can be optimized in many ways, including parsimony, likelihood, and Bayesian approaches. We utilize parsimony, because it was used to reconstruct ancestors for the evolutionary rates analysis in the original study, and it is the most common approach for reconstructing ancestors in morphological phylogenetic analyses of extinct vertebrate taxa. We optimized characters in PAUP 4.0 (Swofford 2003) using the two typical, end-member optimizations in a parsimony framework: accelerated (ACCTRAN) and delayed (DELTRAN) optimization. In cases of homoplasy, accelerated optimization favors losses over convergence and places character changes toward the root of the tree. Delayed optimization, on the other hand, favors convergence over losses and places changes toward the tips of the tree. Both optimizations encompass the same number of total changes but differ only in the way these changes are assigned to branches (and thus what states are given to internal nodes). It is worth noting that other parsimony optimization approaches may also be useful, such as using unambiguous changes only (i.e., considering changes only when ACCTRAN and DELTRAN agree), and these may be preferred by some authors.

Temporal Binning of Reconstructed Ancestors.—The reason for adding a phylogenetic correction (reconstructed ancestors) is so that morphologies implied by a phylogeny, but currently unsampled in the fossil record, can be added to the disparity analysis. In this case, we are concerned only with adding new morphologies (the reconstructed ancestors), and not with extending observed morphologies (terminal taxa) back in time by reference to the first appearance to their sister taxon (see Norell 1992, 1993). In this sense, our methods are a partial correction, not strictly equivalent to phylogenetic diversity corrections that include such lineage extensions. We prefer a partial correction, because reconstructed morphologies are theoretical concepts that are far more difficult to correct for than simple taxon counts, and thus a more conservative approach is warranted. Reconstructed ancestors are theoretical concepts, however, so it is not always straightforward how to best bin these introduced morphologies. Binning ghost lineages in taxonomic diversity studies, which are a rough equivalent to reconstructed ancestors in a disparity study, is straightforward: by definition, a lineage originates when it diverges from its sister taxon. However, the same is not true of morphology. The morphology of a reconstructed ancestor, or a terminal taxon for that matter, is not necessarily present when a lineage initially diverges, or constant throughout the history of the lineage itself.

Thus, the fundamental question is: How should reconstructed ancestors be binned? Two approaches are useful, which constitute somewhat extreme end-members that should provide an envelope around the true tempo of morphological change in a lineage (Fig. 2). The more conservative approach bins each ancestor in the same time period as its first sampled descendant. This approach is the standard in evolutionary rates analysis, which must bin each branch of the tree in order to calculate average rates over time (Wagner 1997; Ruta et al. 2006; Brusatte et al. 2008a). As such, it was used for the evolutionary rates analysis in the original archosaur study (Brusatte et al. 2008a), and is the most obvious method for comparison with that analysis.

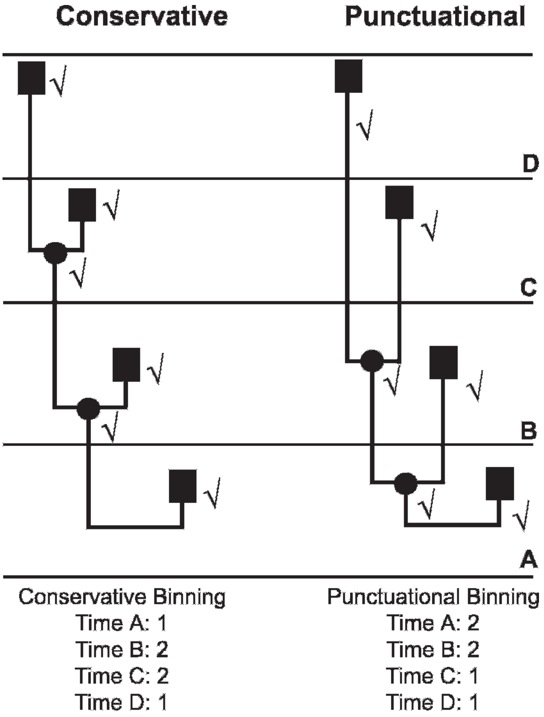

FIGURE 2. A schematic comparison of two approaches for binning hypothetical ancestors in phylogenetically corrected disparity analysis. In both cases the ancestor is binned only in a single time interval and terminal taxa (observed morphologies) are not extended further back in time. The ''conservative'' approach bins ancestors in the same time bin as their first sampled descendant, the same method used in evolutionary rates analysis (sensu Wagner 1997; Ruta et al. 2006; Brusatte et al. 2008a). The ''punctuational'' approach may extend an ancestor further back in time if its sister taxon is older than its oldest sampled descendant.

Additionally, there is a second approach, which is best described as a ''punctuational'' method (Fig. 2). This approach bins each reconstructed ancestor depending on the age of its sister taxon. If the sister taxon is the same age or younger, then the ancestor is binned according to the ''conservative'' method described above (at the time of the first sampled descendant). However, if the sister taxon is older, then the ancestor is binned in the same time period as the sister taxon. In essence, this approach assumes that all evolutionary change in the lineage leading to the reconstructed ancestor, and thus the final morphology of the ancestor, occurred very rapidly, perhaps even associated with the split from its sister taxon as outlined by the punctuated equilibrium or quantum model of evolutionary change (e.g., Simpson 1944; Eldredge and Gould 1972; Gould 2002). Thus, this approach is referred to as the ''punctuational model,'' a term also given to a similar method used recently by Friedman (2009).

We stress that the conservative and punctuational methods are end-members, and are practical methods that deal with the realities of binning taxa, not implied or hypothesized modes of evolutionary change. We also reiterate that we are only binning reconstructed ancestors, and are not concerned with the extension of terminal lineages (largely because terminals can be extended only by the punctuational method and not the conservative method, thus hampering the use of both methods as end-member envelopes). Therefore, terminal lineages are simply binned according to where they are observed in the fossil record. Other authors may prefer to apply other approaches, which may include extending terminal taxa and binning reconstructed ancestors in multiple time bins (not simply the conservative and punctuational extremes). We have not chosen to experiment with these possibilities here, but they may prove useful.

Morphological Disparity.—Two taxon-character data sets were constructed, each of which included data for observed terminal taxa and reconstructed ancestors. Both data sets included the observed scores for all 437 characters for each terminal taxon. To these scores were added, in the first data set, the ancestors as reconstructed by ACCTRAN, and in the second data set, the ancestors as reconstructed by DELTRAN. Thus, each data set included 127 total taxa (64 terminal taxa, 63 reconstructed ancestors), each scored for the full set of 437 characters. From this point on, each data set was analyzed separately, using the methodological protocols described below.

Each data set was used to derive a Euclidean distance matrix, using the freeware MATRIX available from Matthew Wills (personal communication), which quantifies the pairwise dissimilarity between taxa. The distance matrix was then subjected to principal coordinates analysis (PCO), using the freeware program Gingko (Universitat de Barcelona, *http://biodiver.bio.ub.es/ginkgo/*), which condenses information from all 437 characters into a manageable set of axes (in this case, 127, the same as the number of taxa). The first axis represents those characters contributing most to the overall variation among taxa and each additional axis represents features of successively less significance. The various axes also allow each taxon to be plotted within a multidimensional morphospace (a ''taxon defined empirical morphospace'' sensu McGhee 1999). Principal coordinates analysis, as opposed to principal components or other multivariate techniques, is well-suited for disparity analyses because it can better deal with missing data and inapplicable characters, which are common in the present data set (Wills et al. 1994; Lupia 1999).

The PCO analysis returned a set of PCO scores for each taxon, which plot the taxon on each of the 127 morphospace axes. Disparity metrics were calculated using the first 24 PCO axes, which encompass 35% of the total variance, as determined by a substantial break in the slope of the scree plot (see Wills et al. 1994). Taxa were binned by time (see above) and by clade, and disparity metrics were calculated to describe the morphological variability of each bin. Four disparity metrics were calculated: the sum and product of the ranges and variances on the 24 axes (Wills et al. 1994) using the software program RARE (Wills 1998). Multiplicative measures were normalized by taking the 24th root (i.e., the number of PCO axes used [Wills et al. 1994; Stockmeyer Lofgren et al. 2003]). Range measures quantify the entire spread of morphological variation (morphospace size), whereas variance measures denote mean dissimilarity among forms (the spread of taxa in morphospace compared to its center) (see Wills et al. 1994). The former are more sensitive to sample size (and thus sampling bias, because entire intervals might have a better or worse fossil record than other intervals), whereas the latter are less sensitive to sampling biases but may be affected by differences in taxonomic practice, in this case exemplar usage (Wills et al. 1994). Statistical significance between bins was assessed by the overlap or non-overlap of 95% bootstrap confidence intervals, which are calculated by RARE (1000 replications). This test is especially conservative, because it treats the data as two one-sample problems instead of a single two-sample problem, an approach we prefer because it gives stronger confidence to a significant result (S. C. Wang, pers. comm.). Rarefaction curves, also generated by RARE, give an indication of sample-size biases.

Additional Test Cases

The general methods outlined above were also used to analyze two additional data sets, each of which focuses on a separate fossil mammal group that has been the subject of extensive phylogenetic and anatomical study. Protocols for reconstructing ancestors, binning ancestors by time, subjecting the data set to multivariate (PCO) analysis, and calculating disparity metrics are identical to those given above.

The first data set focuses on extant and fossil members of the clade Carnivoramorpha, the large group including dogs and cats as well as the extinct miacids and viverravids. We utilized the cladistic data set of Wesley-Hunt and Flynn (2005), which includes 40 taxa scored for 99 discrete anatomical characters, including those related to dentition and soft tissues. The included taxa span most of the Cenozoic, from the Eocene to the Recent. To obtain a single, resolved tree for ancestral reconstruction, the data set was subject to a heuristic search in PAUP 4.0 (tree bisection and reconnection, 1000 replicates [Swofford 2003]). Eight most parsimonious trees were recovered, the first of which was chosen for the disparity analysis, following protocol in Brusatte et al. (2008a). The phylogenetic corrections added 37 new taxa to the analysis, resulting in 77 total taxa scored for the 99 characters. Taxa were binned into five intervals (Eocene, Oligocene, Miocene, Pliocene, Pleistocene–Recent), and disparity metrics were calculated using the number of PCO axes that described 90% of total variance.

The second data set focuses on euarchontogliran mammals from the Mesozoic and Cenozoic, including glires (rodents and lagomorphs) and their close relatives. We utilized the morphological data matrix of Asher et al. (2005), which includes 68 taxa scored for 228 characters. The phylogeny was reconstructed in TNT (New Technology Search, Ratchet, 68 iterations, TBR Collapsing rule [Goloboff et al. 2003]), using the original data set (which also included molecular data), resulting in 68 most parsimonious trees (13,733 steps). For disparity analysis, the first most parsimonious tree was chosen following Brusatte et al. (2008a). The phylogenetic calibration added 65 additional taxa, resulting in 130 total ingroup taxa scored for the 228 characters. The taxa ranged from the Late Cretaceous to present, and were binned into seven time intervals (Late Cretaceous; Paleocene; Eocene 1: Ypresian to Lutetian; Eocene 2: Bartonian to Priabonian; Oligocene; Miocene 1: Aquitanian to Langhian; and Miocene–Recent: Serravalian to Recent), each of which spans approximately 10 Myr, except the Late Cretaceous bin, which spans 25 Myr (Coniacian to Maastrichtian). Disparity metrics were calculated using the first 22 PCO axes (40% of total variance) according to the significant change in the scree plot.

Additional Approaches.—Our three data sets utilize discrete characters, but the methods outlined above could also be applied to continuous data, such as linear measurements or morphometric shape data. In these cases, it would be necessary to predict what size or shape characterized each hypothetical ancestor. Such optimizations can also be performed in a parsimony setting, and two common methods are linear parsimony (Farris 1970) and squared change parsimony (Maddison 1991). These can be implemented in software such as Mesquite (Maddison and Maddison 2009) and TNT (Goloboff et al. 2003).

Results

Archosauria

Morphospace Plots.—The ACCTRAN and DELTRAN data sets produce similar morphospaces when subjected to PCO analysis (two-dimensional plots presented in Fig. 3). However, there are differences between the phylogenetically corrected morphospaces and

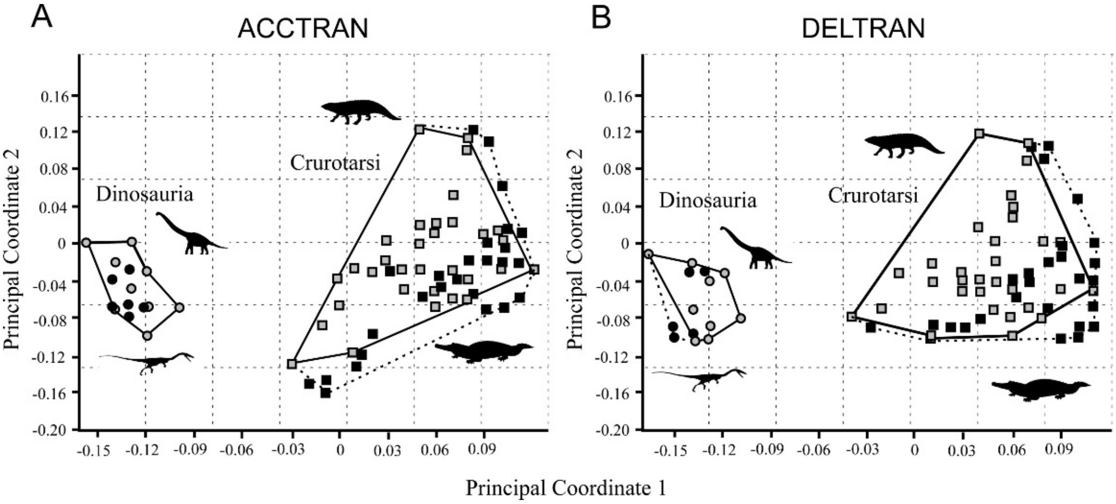

FIGURE 3. Two-dimensional morphospaces, based on the first two principal coordinate axes, for Triassic archosaurs. These morphospaces are based on the phylogenetically corrected disparity analysis reported in the text (A, ancestors reconstructed with accelerated transformation assumption; B, ancestors reconstructed with delayed transformation assumption). For simplicity, only dinosaurs and crurotarsans are illustrated, and archosaurs falling outside of these groups (including pterosaurs and non-dinosaurian dinosauromorphs) are not shown. Additionally, taxa that are essentially overlapping in morphospace are shown only with a single point for clarity. Gray circles (dinosaurs) or squares (crurotarsans) represent observed terminal taxa, and black circles or squares represent reconstructed ancestors. Black lines are convex hulls encompassing the morphospace as defined by observed taxa, and dotted lines indicate extensions in morphospace due to the position of reconstructed ancestors.

the taxic morphospace defined solely by the observed terminal taxa. Most important, the reconstructed ancestors do not uniformly fall within the morphospace envelope defined by the terminals. This is especially true in the crurotarsan region of morphospace (the region containing the ''crocodile-line'' archosaurs), in which reconstructed ancestors are generally placed in a more positive position on PCO axis 1 and a more negative position on PCO axis 2 relative to the observed taxa. This pattern is seen in both ACCTRAN and DELTRAN morphospaces, and serves to expand the morphospaces into the upper and lower right quadrants. Reconstructed dinosaur ancestors are mostly placed in more negative regions on both PCO axes relative to the observed taxa, but the displacement of ancestors from terminals is not as great as within the crurotarsan region of morphospace. Hence, the introduction of ancestors results in a larger overall morphospace, and not an obviously tighter packing within morphospace. In other words, ancestors seem to represent taxa with distinctive body plans that increase the overall range of morphologies. This is an interesting result, because reconstruction of ancestors is based on reference to the terminal taxa, and thus it may be expected that ancestors would largely fall within the morphospace limits of the observed taxa.

Phylogenetic versus Taxic Measures.—A comparison of taxic and phylogenetic curves across the Triassic, the latter calculated with taxa binned according to the ''conservative'' method, are presented in Figures 4 and 5. In general, the phylogenetic metrics are greater, often significantly so, than the taxic measures. This is especially true of the two additive metrics, sums of ranges and variances. However, there are instances in the multiplicative curves where the phylogenetic metrics are less than their taxic counterparts. The marked discrepancy between the two metrics in the Norian, the final time bin, is almost certainly due to an edge effect, because there are no Jurassic taxa in the data set, and thus no Jurassic ancestral morphologies to extend back into the Norian. However, the product of variances curve, with ancestors reconstructed with DELTRAN optimization, exhibits an interesting pattern in which all phylogenetic measures are less than their taxic

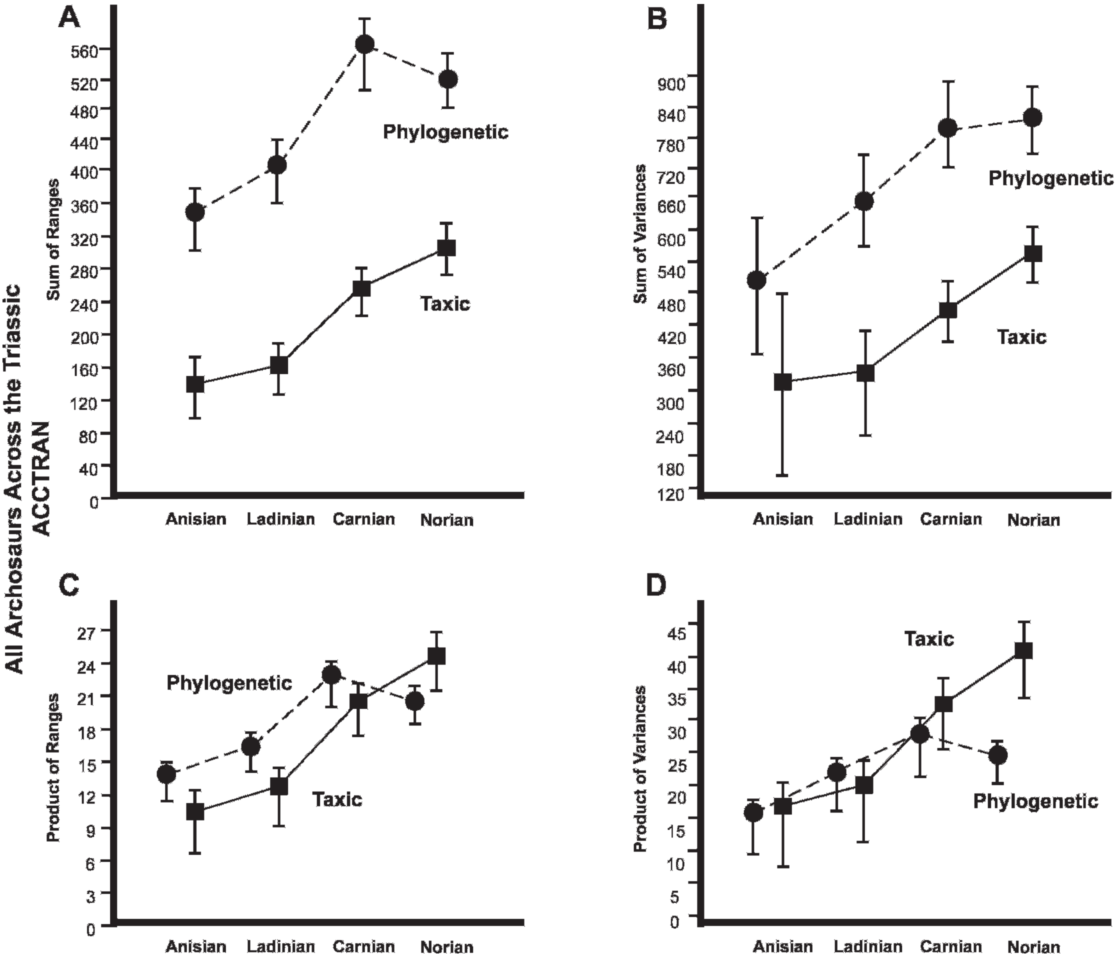

FIGURE 4. Temporal trends in archosaur disparity across the Triassic, measured with four standard metrics (sums and products of ranges and variances on the first 24 PCO axes). Both taxic and phylogenetically corrected trends (with ancestors reconstructed by ACCTRAN and binned by the "conservative" approach) are shown. Error bars are 95% confidence intervals based on bootstrapping (1000 replications), and the non-overlap of bars indicates a statistically significant difference between two time bins. Similarity between the two curves is quantitatively assessed by using first differences correlation (Table 1).

counterparts. Thus, although the phylogenetic corrections mostly ratchet up disparity values (as will almost always happen with range metrics), there are cases in which they can lower them as well.

Most important, however, is that the overall shapes of the taxic and phylogenetic curves are quite similar, which is borne out by correlation of first differences (Table 1). Aside from the Norian edge effect, all curves show a general increase in disparity throughout the Triassic. Rarefaction analysis (not reported, but results are essentially identical to those reported by Brusatte et al. 2008a) indicates that differences between time bins are robust down to a sample size of five for all metrics, which is much smaller than the smallest actual sample (Anisian, $n = 22$). Major increases in taxic disparity are usually also reflected by the phylogenetic disparity metric. Thus, the phylogenetic correction does not automatically lead to amplified disparity values earlier in archosaur history. Despite the extension of unsampled morphologies into earlier time bins, the phylogenetic disparity curves are still discordant with the decreasing trend of evolutionary rates across the Triassic.

Conservative versus Punctuational Binning Approaches.—When taxa are binned according

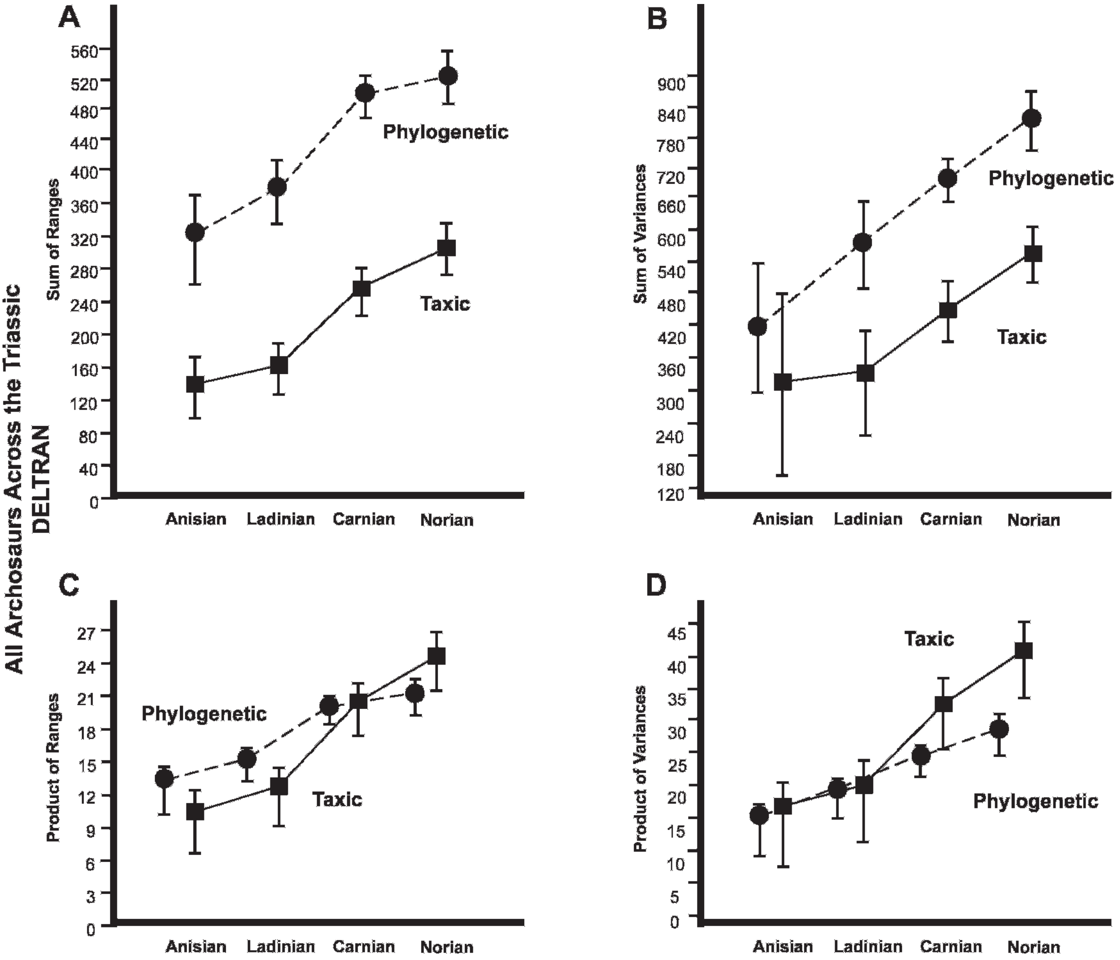

FIGURE 5. Temporal trends in archosaur disparity across the Triassic, measured with four standard metrics (sums and products of ranges and variances on the first 24 PCO axes). Both taxic and phylogenetically corrected trends (with ancestors reconstructed by DELTRAN and binned by the "conservative" approach) are shown. See Figure 4 for additional details of statistical comparisons and confidence intervals.

to the more extreme "punctuational" approach, the disparity of early time intervals (Anisian and Ladinian) is amplified compared to the disparity as calculated with the "conservative" approach (Fig. 6). On the other hand, the disparity within later time intervals (Carnian and Norian) is usually decreased relative to the conservative measures when the punctuational approach is used. Some of this decrease, especially that of the Norian, is likely due to the edge effect described above. However, it is expected that the "punctuational" extension of morphologies into the earliest time interval possible will ratchet up disparity in older time bins at the expense of later time bins. Regardless, the overall shapes of the conservative and punctuational curves are similar: both show a gradual increase in disparity across the Triassic (except for the Norian edge effect) and major increases in one curve are reflected in the other. This is demonstrated quantitatively by correlations of first differences, which indicate that the two binning approaches result in statistically indistinguishable curves (Table 1). Once again, the overall increase in disparity during the Triassic contrasts with the decreasing evolutionary rate trend.

Disparity versus Evolutionary Rate.—An overall summary of phylogenetic disparity (sum of ranges, conservative approach), taxic

TABLE 1. First difference statistical tests comparing various disparity curves (Figs. 4–6, 9, 10) to quantify correlation between them. Comparisons include taxic versus phylogenetic curves (using conservative binning) and conservative versus punctuational binned curves for phylogenetically corrected analyses. Spearman's *r* gives the rank correlation of the two curves in question, with corresponding *p*-value. Because sample sizes are low (each individual curve only depicts 4–7 time intervals) these comparisons include pooled data for the four types of metrics (sums and products of variances and ranges). Some comparisons pool first differences from all four curves constructed with phylogenetic corrections based on a single type of character optimization (all ACCTRAN, all DELTRAN); other comparisons pool only first differences of range or variance curves together (from both ACCTRAN and DELTRAN). In sum, the tests show that taxic and phylogenetic curves are similar to each other, often (but not always) to a statistically significant extent. Conservative and punctuational binning result in a statistically significant correlation in each case examined, indicating that these two end-member binning approaches deliver essentially identical patterns.

Description	Spearman's *r*	*p*-value
Archosaurs: Taxic vs. Phylogenetic		
All ACCTRAN	0.462	0.131
All DELTRAN	0.748	0.005
All range metrics	0.622	0.030
All variance metrics	0.551	0.060
Archosaurs: Conservative vs. Punctuational		
All ACCTRAN	0.797	0.002
All DELTRAN	0.993	0.000
All range metrics	0.951	0.000
All variance metrics	0.811	0.001
Carnivoramorphs: Taxic vs. Phylogenetic		
All DELTRAN	0.891	0.000
All range metrics	0.976	0.000
All variance metrics	0.833	0.010
Euarchontoglirans: Taxic vs. Phylogenetic		
All DELTRAN	0.587	0.003
All range metrics	0.713	0.013
All variance metrics	0.226	0.242

disparity (sum of ranges), and evolutionary rates (from Brusatte et al. 2008a) is presented in Figure 7. As is clear, the phylogenetic correction does not alter the overall trend, evident in the taxic curve, of increasing archosaur disparity across the Triassic. This trend is discordant with the trend of morphological rates over time, so the discrepancy between the two curves still remains despite the phylogenetic correction. In other words, even though both evolutionary rates and disparity measures are now ''phylogenetic'' in nature and more directly comparable, the rapid evolution of discrete characters early in archosaur history still does not appear to have translated into a wide range of morphologies and body plans (and perhaps diets, lifestyles, and other ecological attributes if these are associated with discrete morphological characters).

Taxonomic Comparisons: Dinosaurs versus Crurotarsans.—The crux of this paper has focused on temporal trends in disparity, because phylogenetic corrections are most useful for extending unsampled morphologies backward in time. However, it is also useful to assess whether phylogenetic corrections are useful for taxonomic comparisons. In other words, might phylogenetic corrections improve, or even alter, comparisons of disparity between two or more taxonomic clusters? The observed morphospace patterns described above, in which crurotarsan ancestors are displaced further from the terminal taxa than dinosaur ancestors, suggest that this is true. However, because these are only two-dimensional plots, it is more instructive to analyze the range and variance on the full set of PCO axes.

Brusatte et al. (2008a) compared the disparity of Triassic dinosaurs with that of crurotarsans, the ''crocodile-line'' archosaurs that are found alongside early dinosaurs, converged heavily on dinosaur body plans, and are thought to have been the most important early ''competitors'' of dinosaurs (e.g., Nesbitt 2007). The original disparity study found that crurotarsans were significantly more disparate than dinosaurs no matter the disparity metric used, and that crurotarsans were exploring about twice as much morphospace as dinosaurs during the Triassic. When the same tests are carried out with the phylogenetic corrections nearly the same results are reported (Fig. 8). Crurotarsans are always more disparate than dinosaurs, which rarefaction analysis (not reported, but see Brusatte et al. 2008a for extremely similar rarefaction profiles) indicates is robust to sample-size differences between the groups. However, as is expected with additive metrics, the sums of ranges and variances for each group are ratcheted up. In other words, the overall disparity of each group (both the overall spread of morphospace and the average dissimilarity among forms within

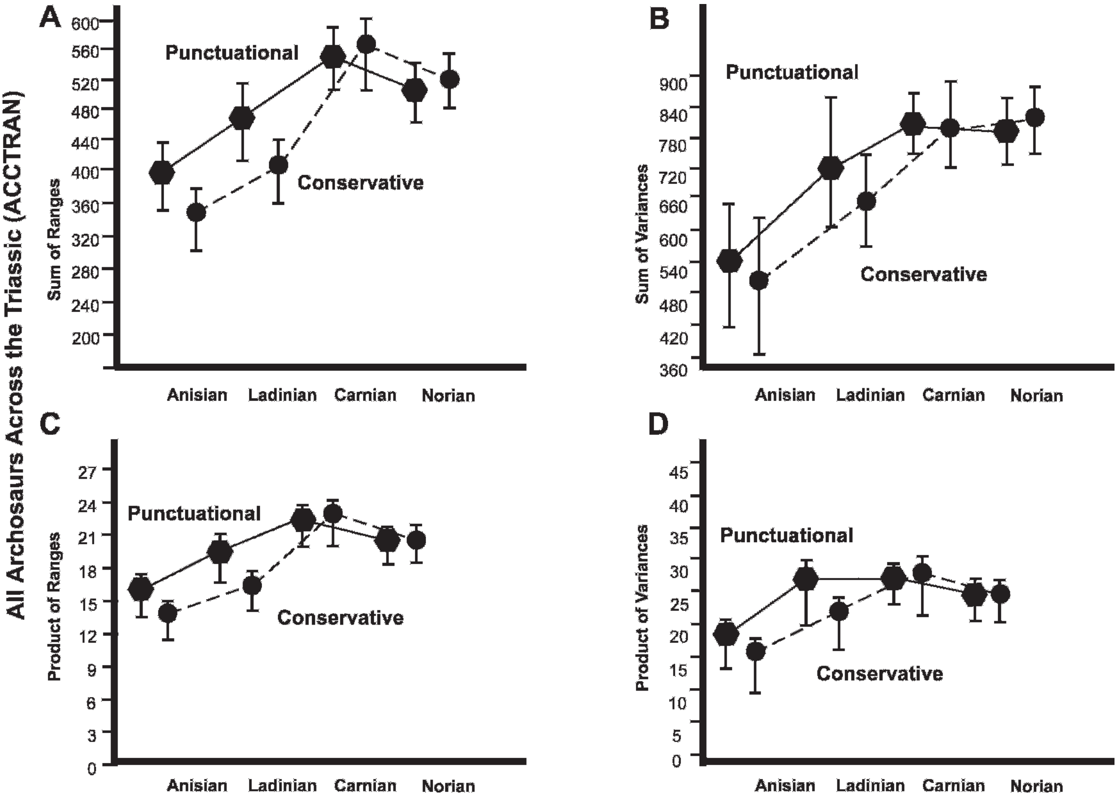

FIGURE 6. Temporal trends in archosaur disparity across the Triassic, measured with four standard metrics (sums and products of ranges and variances on the first 24 PCO axes). Both curves are phylogenetically corrected, but reconstructed ancestors are temporally binned with either the "conservative" or "punctuational" approach (see text). Ancestors reconstructed by ACCTRAN. DELTRAN reconstructions are not shown, but are similar to the ACCTRAN results. See Figure 4 for additional details of statistical comparisons and confidence intervals.

morphospace) is increased. Products of ranges are nearly identical to those reported in the original analysis, whereas products of variances are lower (but not significantly so) than the taxic measures. As this is a multiplicative measure, this pattern is probably due to the influence of very small variance measures on less informative PCO axes. Overall, these quantitative results support the observation in the 2-D morphospace plots that reconstructed ancestors often fall outside of the morphospace limits defined by observed taxa (lead to an increased range), but do not lead to a substantially tighter packing in morphospace (lead to a decreased variance).

Perhaps most interesting, the bootstrap error bars on the sums of variance measures are enormous when the phylogenetic correction is applied, which renders the difference between dinosaurs and crurotarsans nonsignificant. This is true when ancestors are reconstructed with both ACCTRAN and DELTRAN, but is not true of the original taxic analysis, or the other phylogenetically corrected range and variance metrics, which have much tighter error bars. This finding is not easily explained, but at the least it suggests that phylogenetic corrections may, in some cases, give different results than purely taxic analyses (in this case, a nonsignificant and a significant comparison, respectively).

Mammal Data Sets

Carnivoramorpha.—A morphospace with real and hypothetical ancestors is shown in Figure 9A and temporal trends in taxic and phylogenetic disparity are shown in Figure 10. Nearly identical results are obtained when taxa are binned using the conservative

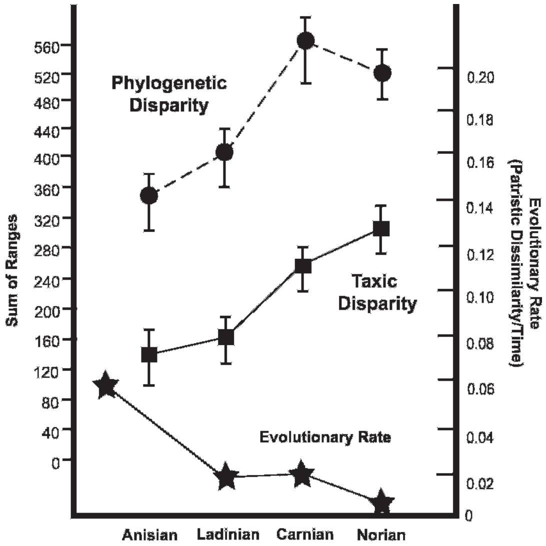

FIGURE 7. A comparison of archosaur taxic disparity, phylogenetically corrected disparity (with ancestors reconstructed by ACCTRAN and binned with the "conservative" approach), and morphological evolutionary rates (characters changing per branch/time duration per branch, with a correction for missing data, sensu Wagner 1997; Ruta et al. 2006, and Brusatte et al. 2008a). Both taxic and phylogenetic disparity increase throughout the Triassic, whereas evolutionary rates generally decrease from an Anisian high to a Norian low (these differences are significant: see Brusatte et al. 2008a). Error bars on disparity measures are 95% confidence intervals based on bootstrapping (1000 replications), and the non-overlap of bars indicates a statistically significant difference between two time bins.

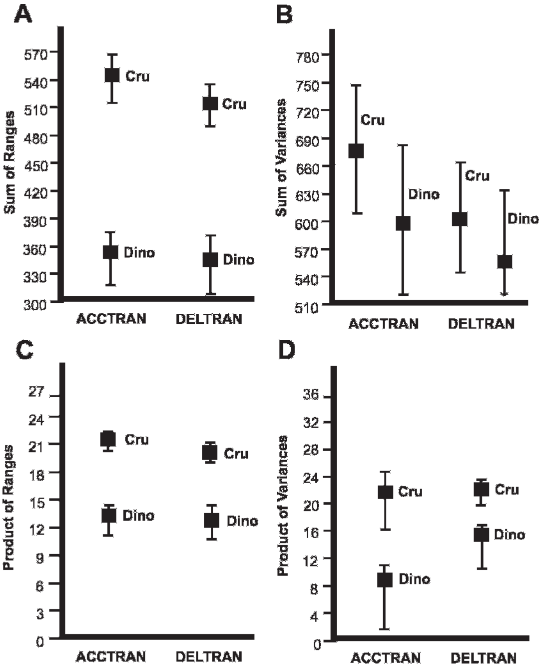

FIGURE 8. Phylogenetically corrected disparity metrics (sums and products of ranges and variances) for Triassic crurotarsans and dinosaurs. Crurotarsan disparity is always higher than dinosaur disparity, and this difference is significant for three of the four metrics. The results are similar to those reported in the original, taxic study by Brusatte et al. 2008a. Error bars are 95% confidence intervals based on bootstrapping (1000 replications), and the non-overlap of bars indicates a statistically significant difference between two time bins.

and punctuational method, as well as when ancestors are reconstructed using ACCTRAN and DELTRAN, and thus only the phylogenetic corrections using DELTRAN and the conservative binning are shown in the figures. The morphospace with hypothetical taxa is expanded relative to that defined by observed taxa only. The phylogenetic measures are, in all cases, greater than the taxic measures for a single time bin, and the discrepancy is greatest for the two additive metrics. Overall, as indicated by first differences correlation, the taxic and phylogenetic curves are significantly similar with each other (Table 1). However, despite their overall similarities, the two curves do exhibit important differences. Most important, the taxic method produces essentially flat curves with no statistically significant disparity differences between time bins. On the other hand, the sum and product of ranges metrics exhibit amplified values in the Eocene, the earliest time bin, as well as depressed values in the Oligocene. Disparity is significantly greater in the Eocene relative to the Oligocene, which is likewise significantly lower relative to the Miocene. The sum and product of variances exhibit similarly shaped taxic and phylogenetic curves, but the phylogenetic corrections indicate that some differences between bins are significant. Interestingly, there are no obvious edge effects that lead to decreased disparity in the final time interval.

These results show that, unlike with the archosaur data set, it is possible for phylogenetic correction to lead to amplified disparity values in earlier time bins, due to the extension of unsampled morphologies back in time. Although first differences indicate statistically significant correlations between the taxic and phylogenetic curves (Table 1), it

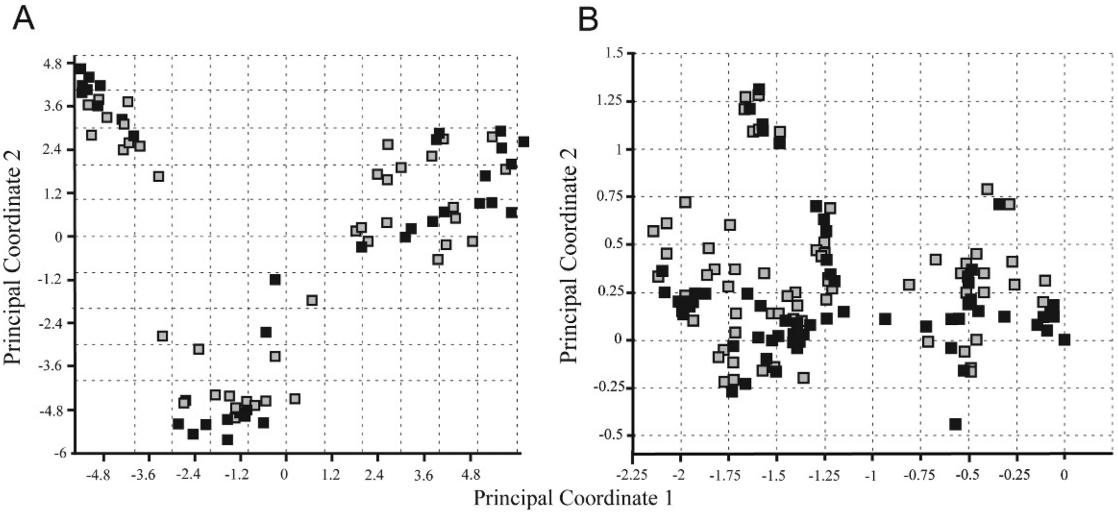

FIGURE 9. Two-dimensional morphospaces, based on the first two principal coordinate axes, for carnivoramorphan mammals (A) and euarchontogliran mammals (B). Only DELTRAN morphospaces are shown, but those that include ancestors reconstructed with ACCTRAN give similar results. Gray squares represent observed terminal taxa and black squares represent reconstructed ancestors. In both examples reconstructed ancestors slightly expand the size of the morphospace as defined by observed taxa.

is clear that the nuances of the two curves differ, and that these differences have interesting macroevolutionary implications for the timing of significant increases and decreases in disparity.

Euarchontogliran Mammals.—A morphospace with real and hypothetical ancestors is shown in Figure 9B and temporal trends in taxic and phylogenetic disparity are shown in Figure 11. For the reasons mentioned above, only the phylogenetic corrections using DELTRAN and conservative binning are shown. The morphospace with hypothetical ancestors is slightly expanded relative to that defined by observed taxa. In general, the phylogenetic curves are similar to the taxic curves, which is borne out by first differences correlation (Table 1; but the variance curves are significantly different from each other). However, the size of the significant jump between the Late Cretaceous and Paleocene is greater (i.e., more significant) when the range metrics are phylogenetically corrected. This is noteworthy, because there is continuing debate regarding the importance of the K/T transition for mammalian evolution (e.g., Alroy 1999; Springer et al. 2003; Bininda Emonds et al. 2007; Wible et al. 2007). Additionally, although nonsignificant, three of the four phylogenetic measures find Paleocene disparity to be greater than Eocene disparity. This is never found with the taxic metrics, and may indicate a flowering of euarchontan body plans during the Paleocene that is not picked up in a direct reading of the fossil record, and thus be further evidence for an important radiation of mammalian biodiversity during this time.

Discussion

Unsampled lineages, and their concomitant morphologies, are an unavoidable reality of many phylogenetic hypotheses. For example, the presence of crocodile-line archosaurs in the Early Triassic predicts that their sister taxon, the bird-line archosaurs, should have also been present at this time. However, bird-line fossils have yet to be observed in the Early Triassic, and therefore the required presence of bird-line morphologies is completely ignored in measures of Early Triassic archosaur disparity (e.g., Brusatte et al. 2008a,b). This is clearly not ideal and results in an incomplete assessment of morphological variety. More pressing, in some cases it may lead to inaccurate and unreasonable measures of disparity. The method we have introduced provides one potential remedy, although

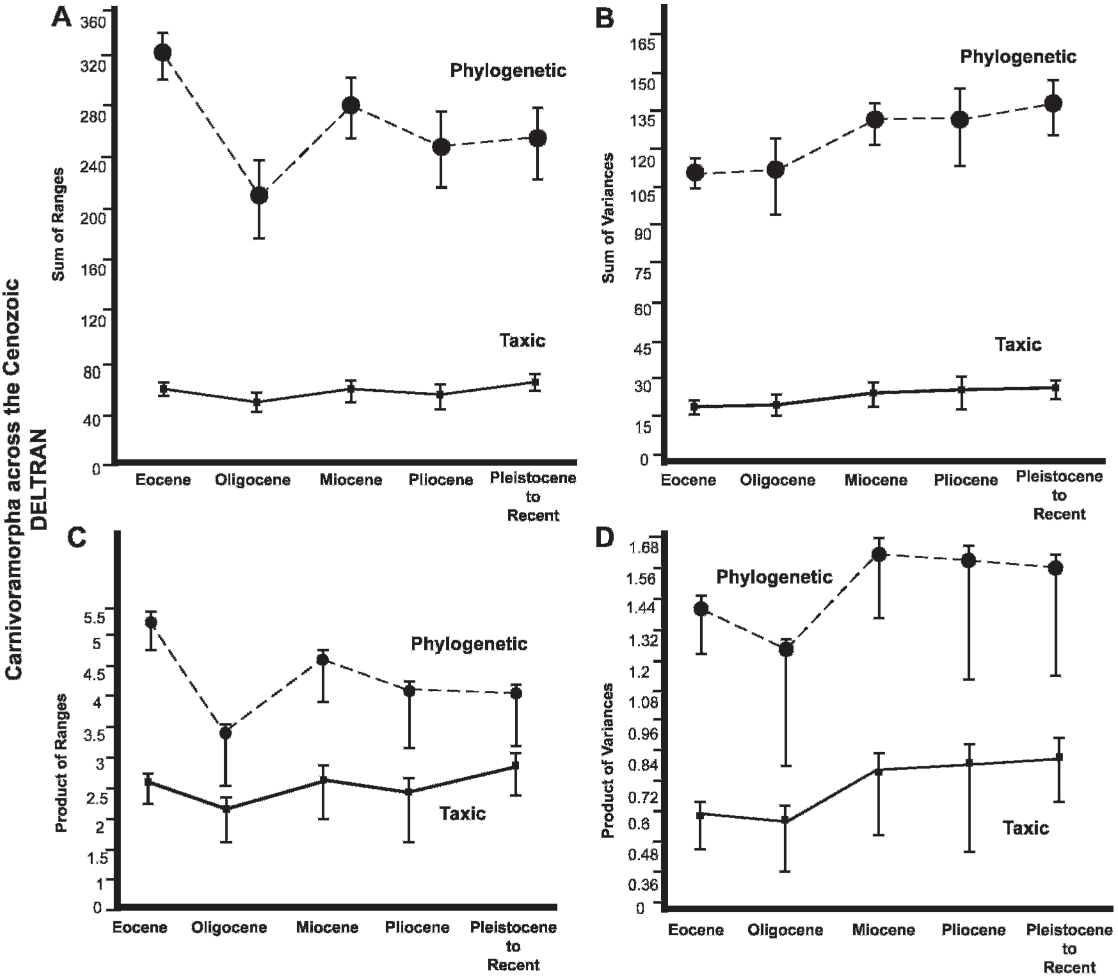

FIGURE 10. Temporal trends in carnivoramorph mammal disparity across the Cenozoic, measured with four standard metrics (sums and products of ranges and variances on the PCO axes). Both taxic and phylogenetically corrected trends (with ancestors reconstructed by DELTRAN and binned by the ''conservative'' approach) are shown. Error bars are 95% confidence intervals based on bootstrapping (1000 replications), and the non-overlap of bars indicates a statistically significant difference between two time bins. Similarity between the two curves is quantitatively assessed using first differences correlation (Table 1).

other remedies are possible and further research is needed to determine those exact scenarios in which taxic disparity is likely to be misleading and phylogenetic corrections are likely to be most useful.

Taxic versus Phylogenetic Disparity Measures

Judging from the test cases presented here, one key finding is that phylogenetic corrections do alter the results of traditional taxic analyses, and these differences are heterogeneous depending on the nature of individual data sets, phylogenetic trees, and disparity comparisons. The most fundamental differences between taxic and phylogenetic measures are seen in temporal disparity curves. In some cases, such as with the Triassic archosaur and euarchontogliran mammal test groups, the phylogenetic curves largely mirror the taxic curves, and the two are usually, but not always, statistically indistinguishable from each other (Table 1). Nonetheless, despite overall congruence between the curves, the magnitude of significant differences between time bins can increase or decrease when disparity is corrected phylogenetically. In the case of euarchontoglirans this has obvious implications for the often-debated

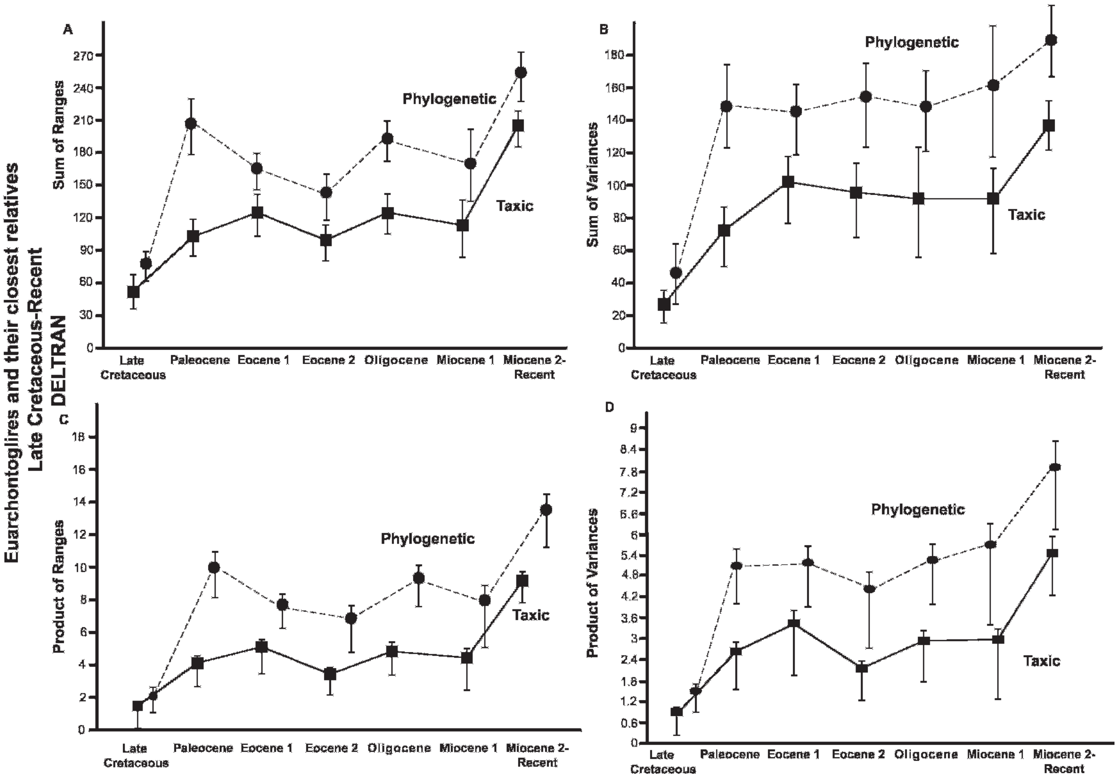

FIGURE 11. Temporal trends in euarchontogliran mammal (Glires and their fossil relatives) disparity from the Late Cretaceous to present, measured with four standard metrics (sums and products of ranges and variances on the PCO axes). Both taxic and phylogenetically corrected trends (with ancestors reconstructed by DELTRAN and binned by the "conservative" approach) are shown. See Figure 4 for additional details of statistical comparisons and confidence intervals.

issue of exactly how much influence the K/T event had on the large-scale patterns of mammalian macroevolution. The much larger increase in disparity between the Cretaceous and Paleocene in the phylogenetically corrected curve may be interpreted as evidence for a significant burst of body plan evolution after the K/T mass extinction, contrary to some suggestions that the K/T event had little effect on mammalian evolution (e.g., Bininda-Emonds et al. 2007). A more profound deviation between the taxic and phylogenetic curves is seen with the carnivoramorph mammal data set. In this case, the disparity is greater in earlier time bins than in later bins after phylogenetic correction, which transforms a static taxic curve into a phylogenetic trend with numerous significant increases and decreases in disparity between intervals. Thus, it is clear that phylogenetic corrections do not simply ratchet up disparity equally in all intervals, but can lead to a relative increase in some intervals (e.g., Eocene) and relative decreases in others (e.g., Oligocene).

Two of our data sets (archosaurs and euarchontoglires) result in taxic and phylogenetic curves that largely mirror each other, whereas the carnivoramorphs exhibit greatly altered phylogenetic curves. This discrepancy is largely due to phylogeny and tree shape. If stratigraphically older taxa are found to be highly derived in the phylogeny, then a number of internal lineages will be dragged further back in time. Similarly, in concert with the age and phylogenetic position of terminal taxa, some tree shapes (balanced versus pectinate and everything in between) may result in more lineages being assigned to a specific time bin than to other bins, just as certain tree shapes are more influenced by a single old fossil in traditional ghost lineage diversity studies (Norell 1993). Consequently, some time bins will now have far more lineages than other bins, as well as more than the number of the observed lineages in the corresponding bin in the taxic analysis, and this increase will likely change the disparity measure. Indeed, this seems to be the case with our sample data sets. Archosaurs and euarchontoglires have phylogenies largely congruent with the stratigraphic ordering of taxa, whereas some old carnivoramorphs are positioned in a derived position, thus dragging a number of internal lineages further back in time and causing an increase in early disparity relative to later time intervals. The important message is that phylogeny and tree shape be carefully considered when undertaking, and especially interpreting, phylogenetic corrections.

Phylogenetic corrections can also result in the expansion of two-dimensional (or multidimensional) morphospace, and the relative amount of expansion may differ between taxonomic groups (e.g., crurotarsans versus dinosaurs). Associated with this, corrections may also change not only the absolute values of disparity metrics, but also the statistical significance of comparisons between taxonomic groups (e.g., the nonsignificant differences between dinosaurs and crurotarsans with the sum of variances measure). These differences are especially important to keep in mind when absolute values of disparity are being considered, as is sometimes the case in morphometric studies (e.g., Anderson 2008). Thus, phylogenetic corrections may make a difference when the goal is to compare disparity between different taxonomic, ecological, or geographic clusters of taxa. One striking result, clearly evident in Figures 4–5 and 10–11, is that corrected disparity measures are much higher than taxic measures. Such a large increase is not normally seen in phylogenetically corrected diversity curves, but these corrections usually add only a small number of ghost lineages and taxa implied by the phylogeny, whereas our phylogenetic disparity corrections essentially double the number of taxa analyzed by including every reconstructed ancestor.

The Use of Phylogenetic Corrections

Keeping in mind the heterogeneous differences between the taxic and phylogenetic curves in our three data sets, phylogenetic corrections may offer a significant improvement in some disparity analyses. A similar finding, based on a theoretical example, was provided by Smith (1994), who argued that phylogenetic methods are superior to purely distance-based approaches, because phenetic methods underestimate the number of characters changing, and hence the disparity between taxa. However, Smith (1994) did not include reconstructed ancestors in disparity analysis, but simply used a patristic distance matrix (calculated on the basis of taxon-taxon differences on a phylogenetic tree) instead of a Euclidean matrix. We appreciate Smith's (1994) general argument but hold that phenetic distances should be used in disparity studies—distantly related taxa that converge on a similar morphology should, ideally, be placed in the same region in morphospace (Gould 1991). Including reconstructed ancestors, but treating these as real taxa that are then compared with others phenetically, is our preferred solution. Smith's (1994) patristic disparity methods are analogous to evolutionary rates analysis (= amount of evolution; see below), which should be kept separate from, but then compared to, purely phenetic disparity analyses (= variety of evolution) to give a more complete picture of morphological evolution.

Our primary goal has been to introduce a set of methods and use them in a select few test cases, and in each case the corrections have altered the results and interpretations of the taxic analyses. This is analogous to the use of phylogenetic corrections in taxonomic diversity studies, which oftentimes drastically alter counts, trends, and patterns based on observed taxa alone. However, these methods have not enjoyed universal acceptance in the paleobiology community. First, the phylogenetic relationships of some groups are woefully understudied, and any phylogenetic corrections, whether for diversity or disparity, rise or fall on a robust phylogeny. Second, several authors have argued that phylogenetic corrections for diversity studies offer only a marginal return for the effort, and in some cases can lead to inaccurate results (e.g., Sepkoski and Kendrick 1993; Wagner 1995, 2000a,b; Lane et al. 2005).

The same qualms may hold true with phylogenetic corrections for disparity analysis. In particular, disparity metrics (especially variance-based metrics) should theoretically be more robust to sampling biases than counts of diversity are (Wills et al. 1994; Ciampaglio et al. 2001), and perhaps phylogenetic corrections may often prove inconsequential. However, despite overall congruence between most of the taxic and phylogenetic curves in our three test cases (Table 1), the various nuanced, heterogeneous differences between these measures—some of which have important and mutually exclusive implications for macroevolution—are prime evidence that phylogenetic corrections do not simply mimic the taxic pattern.

As a next step, it will be important for future studies to assess the performance of phylogenetic corrections when applied to different data sets and to reveal general conditions under which phylogenetic corrections may give a different result than taxic measures. Particularly useful will be sensitivity analyses, which look at the performance of phylogenetic corrections under different sampling regimes and possible errors (e.g., phylogenetic topology, character optimization), and simulations that use hypothetical data to explore the behavior of phylogenetic corrections using different parameters (balanced versus pectinate trees, patchy fossil record versus more complete fossil record, ancestors whose morphologies are inferred well with reference to the terminal taxa versus other possible alternatives). We leave these analyses to future studies, which could be conducted in analogous fashion to Lane et al.'s (2005) simulation of phylogenetic diversity corrections, but on the basis of our empirical results, we suggest that tree shape and the completeness of a group's fossil record are major controls on whether phylogenetic disparity curves will differ from taxic curves.

Because unobserved morphologies are often a requirement of phylogenetic hypotheses, many authors may always want to use phylogenetic corrections in disparity analyses. Keeping in mind the difficulties of introducing unobserved morphologies, however, we stop short of advocating universal application of phylogenetic corrections, but rather suggest that authors carefully consider the reality of the taxonomic group they are studying and the available data. In the case of Triassic archosaurs and the two mammal groups we studied, there are substantial missing lineages early in the history of many of the major clades, as well as entire time periods where sampling is poor or nonexistent. In these cases phylogenetic corrections may be especially useful in helping to extend morphologies into these poorly sampled gaps. In other cases, however, when there are few major gaps in the fossil record of the group in question and when sampling is dense, phylogenetic corrections may not prove as useful. Ultimately, simulations and sensitivity analyses may be useful guides for authors faced with this decision.

If authors choose to use phylogenetic corrections, they should not do so blindly. It is important to experiment with different binning approaches (conservative, punctuational, and perhaps even other approaches such as extending observed morphologies instead of simply introducing reconstructed ancestors, or randomizations sensu Pol and Norell 2006), just as it is common to optimize characters under different assumptions when studying patterns of morphological evolution. Importantly, authors should be aware of edge effects that necessarily follow phylogenetic methods, because phylogenetic corrections only extend lineages further back in time, and thus cannot extend morphologies into the last-sampled time bin (e.g., Wagner 1995). For these and other reasons, we recommend that authors also provide a taxic disparity analysis alongside those corrected with phylogeny. More than anything, we hold that compiling and comparing taxic and phylogenetic disparity methods provides interesting and fruitful insights into morphological evolution. Major macroevolutionary signals, such as disparity crashes at mass extinctions or great increases at evolutionary radiations, can earn

TABLE 2. Pairwise Mann-Whitney *U*-tests assessing differences in the distribution of temporal branch lengths within the four Triassic time bins. Significant differences are in boldface. Averages within the four bins are as follows: Averages: Anisian (0.64381 Ma), Ladinian (1.565 Ma), Carnian (2.302 Ma), Norian (7.734 Ma). The tests indicate that average branch durations increase over time, and are significantly lower during the Anisian than during any other time period.

	Anisian	Ladinian	Carnian	Norian
Anisian	X	**$U = 73$; $p = 0.000049$**	**$U = 144$; $p = 0.000011$**	**$U = 14$; $p < 0.00000001$**
Ladinian	X	X	$U = 461.5$; $p = 0.4801$	**$U = 69$; $p = 0.00000019$**
Carnian	X	X	X	**$U = 216$; $p = 0.00000044$**
Norian	X	X	X	X

greater confidence if recovered by both methods. If different curves are recovered, however, this is a strong indicator that the observed fossil record is incomplete, and any time periods (or clades) with discordant taxic and phylogenetic patterns should be targeted for future research.

Triassic Archosaurs: Disparity versus Rates

One situation in which authors should always use phylogenetic disparity corrections is when disparity is being compared with the rate of morphological character evolution. The most common rate metric, characters changing per branch of a phylogeny divided by the time duration of that branch, is phylogenetic in nature, not only because it relies on character optimization but also because internal branches (hypothetical ancestral branches) are included in the calculations (Wagner 1997; Ruta et al. 2006; Brusatte et al. 2008a). Previous comparisons of disparity and rate have contrasted a taxic estimate with a phylogenetic one, which is not ideal. One of the main goals of the current study was to reassess the data set of Brusatte et al. (2008a), to determine whether their noted discrepancy between rates and disparity in archosaurs was simply due to such an improper comparison.

Our revised analyses indicate that the gradual increase in archosaur disparity across the Triassic, observed originally with taxic data, is a robust pattern. When the Brusatte et al. (2008a) data set is phylogenetically corrected and disparity is recalculated, the same general trend (gradual Triassic increase), as well as the relative magnitude of increases between time bins, reported in the original study is recovered. The extension of morphologies into more poorly sampled older time bins does not ratchet up Middle Triassic disparity values at the expense of Late Triassic values. Therefore, the discrepancy between disparity and evolutionary rates remains.

A discordant relationship between rate of character change and morphological variety may be unexpected under some models of evolutionary change. For instance, simulations performed by Foote (1996: Fig. 4.3) found that disparity plateaus once step size (here equivalent to morphological rate) slows down. Archosaurs, on the other hand, exhibit a continual increase in disparity for tens of millions of years even after rates slow down, and plateau, from their Middle Triassic peak (Brusatte et al. 2008a). What, then, could explain the divergent relationship between archosaur disparity and rates? Brusatte et al. (2008a) noted that the raw number of characters changing—the number of characters changing per branch regardless of temporal branch length—remains essentially constant throughout the Triassic. However, the Triassic rate decrease is recovered when time enters into the equation. Because time is the denominator in the rate equation, shorter temporal branch lengths earlier in the Triassic, and a pattern of increasingly longer average branch lengths over time, may explain the gradual decrease in rates. Indeed, when the temporal duration of branches are compared by time interval, a statistically significant pattern of increasing lengths over time is apparent (Table 2).

Therefore, it is likely that temporal branch lengths are primarily driving the rate trend in Triassic archosaurs. The question now becomes whether significant differences in branch lengths over time are a real phenomenon or a whether they result from a bias in the fossil record. It could be argued that branch durations should be approximately equal, regardless of when in the Triassic that branch was evolving, and differences may simply be due to the fact that Middle Triassic stages are shorter than those in the Late Triassic (because stage length plays a major role in assigning temporal durations to branches). Alternatively, it could be argued that the Middle Triassic stages are genuinely shorter than stages later in the Triassic, and that the increasing branch length trend simply reflects the reality of the Triassic time scale. Distinguishing between these alternatives is difficult at present, but further study is clearly needed, especially as new Triassic archosaurs are discovered and the Triassic time scale is further refined (e.g., Irmis and Mundil 2008).

Disparity and Rates: Congruence or Discordance?

If the discordance between disparity and evolutionary rates in Triassic archosaurs is a real phenomenon, is this normal or abnormal? Surprisingly, few comparisons of rates and disparity have been made (Wagner 1997; Brusatte et al. 2008a), so it is difficult to gauge common patterns. Null models of constant morphological step size over time (or models of decreasing step size over time) may not predict this pattern, but evolution may progress in such a way that step size increases relative to raw rate of change (Foote 1996). This is one evolutionary model that would explain a rate-disparity discrepancy. Interestingly, recent authors have begun to look more closely at the relationship between rates and disparity. Adams et al. (2009) reported a similar discrepancy: in salamanders, there is no significant association of disparity with rates of shape evolution. Theoretical work also has suggested several ways in which these two measures can be decoupled, mostly because of nuances of the age of clades, the temporal duration of branches, tree shape, and the distribution of temporal branch lengths on a tree (O'Meara et al. 2006; Ricklefs 2006; Sidlauskas 2007). Similarly, other studies indicate that if morphological evolution tends toward an adaptive peak or another constraint, then rates of morphological change can be high but produce very little change in overall disparity, as organismal morphology fluctuates around this peak (e.g., Sidlauskas 2008; Collar et al. 2009). Thus, not only is discordance between rates and disparity possible, but it may actually be an expectation of some evolutionary processes.

Conclusions

Traditional morphological disparity analyses are taxic in nature and measure only the variability of observed taxa. Phylogenetic corrections, which can extend unsampled morphologies implied by phylogeny into poorly sampled time bins, may be useful when there are large gaps in the fossil record of the group being analyzed. Application of phylogenetic corrections often does alter the results of traditional taxic analyses, and these differences are heterogeneous depending on the nature of individual data sets, the shape of the utilized phylogenetic trees, and disparity comparisons. Our three independent data sets indicate that, in some cases, phylogenetic corrections ratchet up disparity in earlier time intervals relative to later intervals and lead to an expansion of morphospace size. However, for Triassic archosaurs, phylogenetic corrections result in a curve similar to that of the taxic analysis, and the discordance between disparity and evolutionary rates noted in a previous study (Brusatte et al. 2008a) still stands.

Acknowledgments

We thank two anonymous reviewers and the editor (M. Carrano) for helpful comments that helped tighten our arguments, and R. Benson, M. Benton, J. Flynn, B. Kraatz, G. Lloyd, J. Louys, J. Meng, M. Ruta, M. Siddall, and S. Wang for discussion. S.L.B. thanks his coauthors on the original archosaur paper for their help, as well as numerous curators (listed in Brusatte et al. 2010a) for access to archosaur specimens in their care. S.L.B. is supported by a National Science Foundation Graduate Research Fellowship, and the original archosaur disparity project was funded by the Marshall Scholarship for study in the U.K., the Paleontological Society, and the University of Bristol Bob Savage Memorial Fund. S.M. is supported by a National Science Foundation Graduate Research Fellowship and H.-Y.Y. by Columbia University. The Division of Paleontology and Richard Gilder Graduate School at the American Museum of Natural History, and the Department of Earth and Environmental Sciences at Columbia University, also provided support for this project. Coauthors focused on different data sets: archosaurs (S.L.B.), carnivoramorphs (S.M.), and euarchontoglirans (H.-Y.Y.).

Literature Cited

Adams, D. C., C. M. Berns, K. H. Kozak, and J. J. Wiens. 2009. Are rates of species diversification correlated with rates of morphological evolution? Proceedings of the Royal Society of London B 276:2729–2738.

Alroy, J. 1999. The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. Systematic Biology 48:107–118.

Alroy, J., et al. 2008. Phanerozoic trends in the global diversity of marine invertebrates. Science 321:97–100.

Anderson, P. S. L. 2008. Shape variation between arthrodire morphotypes indicates possible feeding niches. Journal of Vertebrate Paleontology 28:961–969.

Asher, R. J., J. Meng, J. R. Wible, M. C. McKenna, G. W. Rougier, D. Dashzeveg, and M. J. Novacek. 2005. Stem Lagomorpha and the antiquity of Glires. Science 307:1091–1094.

Barrett, P. M., A. J. McGowan, and V. Page. 2009. Dinosaur diversity and the rock record. Proceedings of the Royal Society of London B 276:2667–2674.

Benton, M. J. 2005. Vertebrate palaeontology, 3d ed. Blackwell, Oxford.

Bininda-Emonds, O. R. P., M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. Nature 446:507–512.

Brusatte, S. L. 2007. The higher-level phylogeny of Archosauria (Tetrapoda: Diapsida). M.Sc. thesis. University of Bristol, Bristol, U.K.

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

———. 2008b. The first 50 myr of dinosaur evolution: macroevolutionary pattern and morphological disparity. Biology Letters 4:733–736.

Brusatte, S. L., M. J. Benton, J. Desojo, and M. C. Langer. 2010a. The higher-level phylogeny of Archosauria (Tetrapoda: Diapsida). Journal of Systematic Palaeontology 8:3–47.

Brusatte, S. L., S. J. Nesbitt, R. B. Irmis, R. J. Butler, M. J. Benton, and M. A. Norell. 2010b. The origin and early radiation of dinosaurs. Earth-Science Reviews 101:68–100.

Ciampaglio, C. N., M. Kemp, and D. W. McShea. 2001. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. Paleobiology 27:695–715.

Collar, D. C., B. C. O'Meara, P. C. Wainwright, and T. J. Near. 2009. Piscivory limits diversification of feeding morphology in centrarchid fishes. Evolution 63:1557–1573.

Eldredge, N., and S. J. Gould. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Pp. 82–115 *in* T. J. M. Schopf, ed. Models in paleobiology. Freeman, Cooper, San Francisco.

Erwin, D. H. 2007. Disparity: morphological pattern and developmental complexity. Palaeontology 50:57–73.

Farris, J. 1970. Methods for computing Wagner trees. Systematic Zoology 19:83–92.

Fisher, D. C. 1982. Phylogenetic and macroevolutionary patterns within the Xiphosurida. Proceedings of the Third North American Paleontological Convention 1:175–180.

Foote, M. 1993. Discordance and concordance between morphological and taxonomic diversity. Paleobiology 19:185–204.

———.1996. Models of morphological diversification. Pp. 62–86 *in* D. Jablonski, D. H. Erwin, and J. H. Lipps, eds. Evolutionary Paleobiology. University of Chicago Press, Chicago.

———. 1997. The evolution of morphological diversity. Annual Review of Ecology and Systematics 28:129–152.

Friedman, M. 2009. Ecomorphological selectivity among marine teleost fishes during the end-Cretaceous extinction. Proceedings of the National Academy of Sciences USA 106:5218–5223.

Goloboff, P. A., J. S. Farris, and K. C. Nixon. 2003. T.N.T.: tree analysis using new technology, Version 1.0. *www.zmuc.dk/public/phylogeny*

Gould, S. J. 1991. The disparity of the Burgess shale arthropod fauna and the limits of cladistic analysis: why we must strive to quantify morphospace. Paleobiology 17:411–423.

———. 2002. The structure of evolutionary theory. Harvard University Press, Cambridge.

Irmis, R., and R. Mundil. 2008. New age constraints from the Chinle Formation revise global comparisons of Late Triassic vertebrate assemblages. Journal of Vertebrate Paleontology 28(Suppl. to No. 3):95A.

Lane, A., C. M. Janis, and J. J. Sepkoski Jr. 2005. Estimating paleodiversities: a test of the taxic and phylogenetic methods. Paleobiology 31:21–34.

Langer, M. C., M. D. Ezcurra, J. S. Bittencourt, and F. E. Novas. 2010. The origin and early evolution of dinosaurs. Biological Reviews 85:55–110.

Lloyd, G. T., K. E. Davis, D. Pisani, J. E. Tarver, M. Ruta, M. Sakamoto, D. W. E. Hone, R. Jennings, and M. J. Benton. 2008. Dinosaurs and the Cretaceous Terrestrial Revolution. Proceedings of the Royal Society, Series B 275:2483–2490.

Lupia, R. 1999. Discordant morphological disparity and taxonomic diversity during the Cretaceous angiosperm radiation: North American pollen record. Paleobiology 25:1–28.

Maddison, W. P. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. Systematic Zoology 40:304–314.

Maddison, W. P., and D. R. Maddison. 2009. Mesquite: a modular system for evolutionary analysis, Version 2.72. *http://mesquiteproject.org*

McGhee, G. R., Jr. 1999. Theoretical morphology. The concept and its applications. Columbia University Press, New York.

Nesbitt, S. J. 2007. The anatomy of *Effigia okeeffeae* (Archosauria, Suchia), theropod- like convergence, and the distribution of related taxa. Bulletin of the American Museum of Natural History 302:1–84.

Norell, M. A. 1992. Taxic origin and temporal diversity: the effect of phylogeny. Pp. 88–118 *in* M. J. Novacek and Q. D. Wheeler, eds. Extinction and phylogeny. Columbia University Press, New York.

———. 1993. Tree-based approaches to understanding history: comments on ranks, rules, and the quality of the fossil record. American Journal of Science 293A:407–417.

Norell, M. A., and M. J. Novacek. 1992. The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. Science 255:1690–1693.

O'Meara, B. C., C. Ane, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

Paul, C. R. C. 1982. The adequacy of the fossil record. Systematics Association Special Volume 21:75–117.

Pol, D., and M. A. Norell. 2006. Uncertainty in the age of fossils and the stratigraphic fit to phylogenies. Systematic Biology 55:512–521.

Raup, D. M. 1972. Taxonomic diversity during the Phanaerozoic. Science 231:1065–1071.

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

Ricklefs, R. E. 2006. Time, species, and the generation of trait variance in clades. Systematic Biology 55:151–159.

Sepkoski, J. J., Jr. 1984. A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions. Paleobiology 10:246–267.

Sepkoski, J. J., and D. C. Kendrick. 1993. Numerical experiments with model monophyletic and paraphyletic taxa. Paleobiology 19:168–184.

Sidlauskas, B. 2007. Testing for unequal rates of morphological diversification in the absence of a detailed phylogeny: a case study from characiform fishes. Evolution 61:299–316.

———. 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. Evolution 62:3135–3156.

Simpson, G. G. 1944. Tempo and mode in evolution. Columbia University Press, New York.

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

Springer, M. S., W. J. Murphy, E. Elzirik, and S. J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proceedings of the National Academy of Sciences USA 100:1056–1061.

Stockmeyer Lofgren, A., R. E. Plotnick, and P. J. Wagner. 2003. Morphological diversity of Carboniferous arthropods and insights on disparity patterns through the Phanerozoic. Paleobiology 29:349–368.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer, Sunderland, Mass.

Wagner, P. J. 1995. Stratigraphic tests of cladistic hypotheses. Paleobiology 21:153–178.

———. 1997. Patterns of morphological diversification among the Rostroconchia. Paleobiology 23:115–145.

———. 2000a. Phylogenetic analyses and the fossil record: tests and inferences, hypotheses and models. *In* D. H. Erwin and S. L. Wing, eds. Deep time: *Paleobiology*'s perspective. Paleobiology 26(Suppl. to No. 4):341–371.

———. 2000b. The quality of the fossil record and the accuracy of phylogenetic inferences about sampling and diversity. Systematic Biology 49:65–86.

Wesley-Hunt, G. D. 2005. The morphological diversification of carnivores in North America. Paleobiology 31:35–55.

Wesley-Hunt, G. D., and J. J. Flynn. 2005. Phylogeny of the Carnivora: Basal relationships among the Carnivoramorphans, and assessment of the position of 'Miacoidea' relative to Carnivora. Journal of Systematic Palaeontology 3:1–28.

Wible, J. R., G. W. Rougier, M. J. Novacek, and R. J. Asher. 2007. Cretaceous eutherians and Laurasian origin for placental mammals near the K/T boundary. Nature 447:1003–1006.

Wills, M. A. 1998. Crustacean disparity through the Phanerozoic: comparing morphological and stratigraphic data. Biological Journal of the Linnean Society 65:455–500.

Wills, M. A., D. E. G. Briggs, and R. A. Fortey. 1994. Disparity as an evolutionary index: a comparison of Cambrian and Recent arthropods. Paleobiology 20:93–131.

Young, M. T., S. L. Brusatte, M. Ruta, and M. B. Andrade. 2010. The evolution of Metriorhynchoidea (Mesoeucrocodylia: Thalattosuchia): an integrated approach using geometric morphometrics, analysis of disparity, and biomechanics. Zoological Journal of the Linnean Society 158:801–859.