

MODELS AND METHODS FOR ANALYZING PHENOTYPIC EVOLUTION IN LINEAGES AND CLADES

GENE HUNT

and

MATTHEW T. CARRANO

Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013-7012.

Abstract.—In this chapter we discuss methods for analyzing continuous traits, with an emphasis on those approaches that rely on explicit statistical models of evolution and incorporate genealogical information (ancestor–descendant or phylogenetic relationships). After discussing the roles of models and genealogy in evolutionary inference, we summarize the properties of commonly used models including random walks (Brownian motion), directional evolution, and stasis. These models can be used to devise null-hypothesis tests about evolutionary patterns, but it is often better to fit and compare models equally using information criteria and related approaches. We apply these methods to a published data set of dental measurements in a sequence of ancestor–descendant populations in the early primate *Cantius*, with the particular goal of determining the best-supported mode of evolutionary change in this lineage. We also assess a series of questions about the evolution of femoral dimensions in several clades of dinosaurs, including testing for a trend of increasing body size (Cope’s Rule), testing for correlations among characters, and reconstructing ancestral states. Finally, we list briefly some additional models, approaches, and issues that arise in genealogically informed analyses of phenotypic evolution.

PRELIMINARIES: SCOPE AND STRATEGY

THE FOSSIL record is a vast archive of lineages and their phenotypes, arrayed through geological time. Paleontologists seek to document the patterns of phenotypic evolution revealed by this record, and, when possible, to link these patterns to drivers and processes. In this chapter, we review some of the analytical methods available for understanding phenotypic evolution. As this topic is potentially enormous, we have chosen to narrow it to a more manageable size. First, we limit the discussion to phenotypic traits that vary continuously, an entirely pragmatic restriction of scope. In addition, we focus almost entirely on methods that (i) specify explicit statistical models of trait evolution and (ii) incorporate genealogy in the form of hypothesized ancestor–descendant or phylogenetic relationships. These restrictions reflect our view that these methods are often the most powerful and flexible means for testing evolutionary hypotheses. Moreover, whereas non-genealogical methods are already well described and known to paleontologists (e.g., McKinney, 1990;

Foote, 1991; McShea, 1994; Roy and Foote, 1997), the analytical tools appropriate for analyses that are model-based and genealogically-informed are less widely appreciated.

After an overview of the importance of genealogy and models, we will devote the bulk of this chapter to demonstrating the methods as applied to ancestor–descendant and phylogenetic data. In doing so, we will rely heavily on two published data sets: dental measurements in an evolutionary sequence of the early primate *Cantius* (Clyde and Gingerich, 1994), and femoral dimensions in several dinosaur clades (Carrano, 2006).

All analyses in this chapter were performed using the R statistical software package (R Development Core Team, 2009), which is freely available for all popular computer operating systems. In addition to its availability, R has many powerful statistical methods built-in, and is a fully functional programming language for analysis and graphics (all figures in this chapter are nearly unmodified R output). Perhaps most importantly for the present work, R is extendible in that users can create modules, called packages, to perform functions not already implemented. We make heavy use of the

packages *paleoTS* (Hunt, 2008c) for analyzing ancestor–descendant series, and *ape* (Paradis et al., 2004) and *geiger* (Harmon et al., 2009) for various phylogenetic analyses. These packages need to be installed and then loaded into the R session prior to running the analyses in this chapter (e.g., `library(ape)` loads the *ape* package once it has been installed). Other packages are relevant to this chapter; readers are referred to the list of packages related to phylogenetic and comparative methods maintained on the R website (<http://cran.r-project.org/web/views/Phylogenetics.html>).

In the text of this chapter, we include the main commands for performing R analyses in a *fixed width font*. The online supplement, in addition to covering a few topics too technical to warrant inclusion in the main text, provides an R workspace file with the data analyzed in this chapter.

THE IMPORTANCE OF GENEALOGY

Whether focused on individual species or on the broad sweep of life, evolutionary inference requires some information about the genealogical relationships among the morphological units being studied. At a coarse level, taxonomy delimits groups of species that (we hope) are more closely related to each other than they are to all outgroup taxa. For paleontologists, more

detailed genealogy comes in two forms: sequences of populations construed to form a chain of ancestor–descendant relationships (Fig. 1.1), or hypotheses that link populations through sister-group relationships into larger clades (Fig. 1.2). Both of the panels in Figure 1 represent a richer source of genealogical information relative to traditional taxonomy, and this richness demands its own set of analytical methods.

In Figure 1, the lines represent the pathways of accumulated evolutionary change, and the points are snapshots of evolving populations or species. We assume that some aspect of phenotype has been recorded for these populations, and the goal is to learn something about how one or more traits evolve. The detailed goals will depend on the specific research agenda, but can include questions such as: What is the general pattern, or mode, of evolution? What are the best estimates for the phenotypes of unobserved ancestors? How fast does a character evolve? Are changes in one character correlated with changes in another, or with changes in the environment?

Under most notions of how phenotypes evolve, changes accrue—the phenotypic state at one point in time is equal to the ancestral condition plus all subsequent changes leading to the focal population. When evolution builds from pre-existing states in this manner, observations across genealogical units will not be

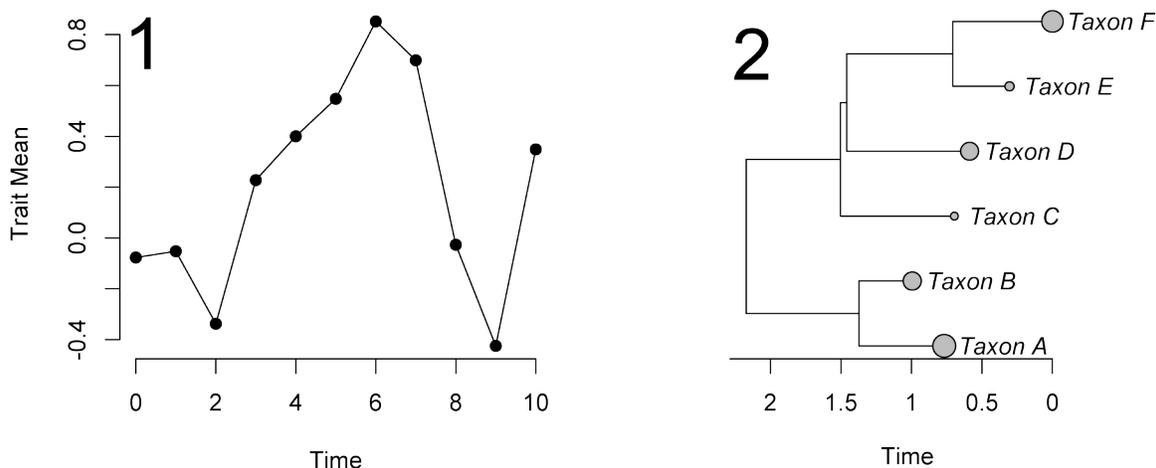


FIGURE 1.—Two kinds of genealogical information in paleontological research. 1, populations forming an ancestor–descendant sequence. 2, populations related through phylogeny. Values of a phenotypic trait are represented by the vertical axis in 1 and by the size of the symbol in 2. Horizontal axes in both represent time, measured forward from the oldest population in 1, and measured backwards from the youngest taxon in 2.

independent. Nearby points in ancestor–descendant sequences share more of their history compared to more distant pairs, and the same is true of closely related species such as Taxa E and F in Figure 1.2. This creates an analytical difficulty because most statistical approaches assume observations are independent, and they can mislead when applied naively to data such as those shown in Figure 1. Faced with dependence among observations, two basic strategies can be used. First, one can construct a new set of observations from the data that are independent. Alternatively, one can accept that the observations are not independent, and account for this statistically. All of the methods that we describe in this chapter rely on one of these two strategies. In order to work correctly, both approaches require assumptions about how evolution proceeds; that is, they assume a particular evolutionary model. Thus it is necessary to first understand common models of phenotypic change.

MODELS OF PHENOTYPIC EVOLUTION

Paleontologists commonly measure features that vary continuously, including sizes of organisms and their parts, and shape metrics such as those derived from geometric morphometrics (Webster and Sheets, this volume). Measurements that reflect size, such as lengths and areas, should usually be log-transformed so they are considered on a proportional scale (see, for example, Foote, 1991; Gingerich, 2000). This is especially true when values span orders of magnitude because proportional differences will be more comparable across scales (e.g., a length increase of 1 cm matters much more to a mouse than to an elephant, but a 10% increase is similar for both). Meristic (count) variables are restricted to integer values and thus are not truly continuous, but they can be reasonably approximated as such as long as the range of realized values is not too small. Discrete characters with few states, such as the presence or absence of structures, arise frequently in studies of phylogenetic relationships and morphological disparity. We will not cover such traits here; see instead Wagner and Marcot (this volume) for a treatment of these traits in the context of phylogenetic analysis. Several other works review some of the important model-based comparative methods for discrete traits (Pagel, 1994; Felsenstein, 2004).

In this chapter, we focus on approaches that model evolutionary outcomes using just a few parameters. Although the models we discuss can be related to microevolutionary scenarios, we do not pursue this avenue extensively here. Except for truly exceptional circumstances (e.g., Hunt et al., 2008), it is usually very difficult to infer specific microevolutionary processes from fossil data. Any paleontological pattern is consistent with a myriad of combinations of selection, drift, mutation and gene flow, and we generally lack the information required to discriminate among possibilities. As a result, it is almost always more profitable to use simple models of evolution to extract meaningful aspects of evolutionary divergence, such as directionality and rates of change, without committing to any specific microevolutionary scenario. We discuss next the three models most commonly considered in paleontological studies: directional evolution, random walks, and stasis.

Directional Evolution and Random Walks.— Directional evolution and random walks are related models that can be derived from a simple evolutionary scenario. Time takes place in discrete increments, which can be equated with biological generations or some other convenient unit such as years. During each increment an evolutionary transition, or “step,” occurs. These steps need not be identical, but instead are drawn at random from a distribution of possible evolutionary changes (Fig. 2). It turns out that only the mean and variance of this distribution matter for long-term evolution. When the mean of the step distribution is positive, increases will outgain decreases, and, on average, trait values will trend upward over time (Fig. 2, middle row). Similarly, a negative mean step produces decreasing trait values. Whenever the mean step is non-zero, we will refer to this model simply as *directional evolution*.

If the mean of the step distribution is set to zero, total trait increases equal decreases, and no trend is expected (Fig. 2, top and bottom rows). This model is sometimes called an unbiased or symmetric random walk. Here we’ll omit the modifiers and refer to this model as a *random walk*. Setting the mean step to zero leaves just one parameter in a random walk: the variance of the step distribution, sometimes called the step variance. If this step variance is high, we will see large increases and decreases in the evolving trait, and correspondingly volatile evolutionary trajectories

(Fig. 2, bottom row). If this variance is more modest, trait changes will be correspondingly smaller (Fig. 2, top row).

If we allow the step size to get smaller and smaller while the time increment between successive steps similarly narrows, in the limit this scenario becomes a diffusion or *Brownian motion (BM)* process. Just like the general random walk, this diffusion process

has two parameters: a mean value that determines the trend (if present), and a variance that determines the variation in outcomes. These are usually referred to as the instantaneous mean (or drift coefficient) and instantaneous variance (or BM coefficient), respectively. As discrete-time and continuous-time versions of the same process, random walks and BM are alike in all properties that matter for the purposes of this chapter.

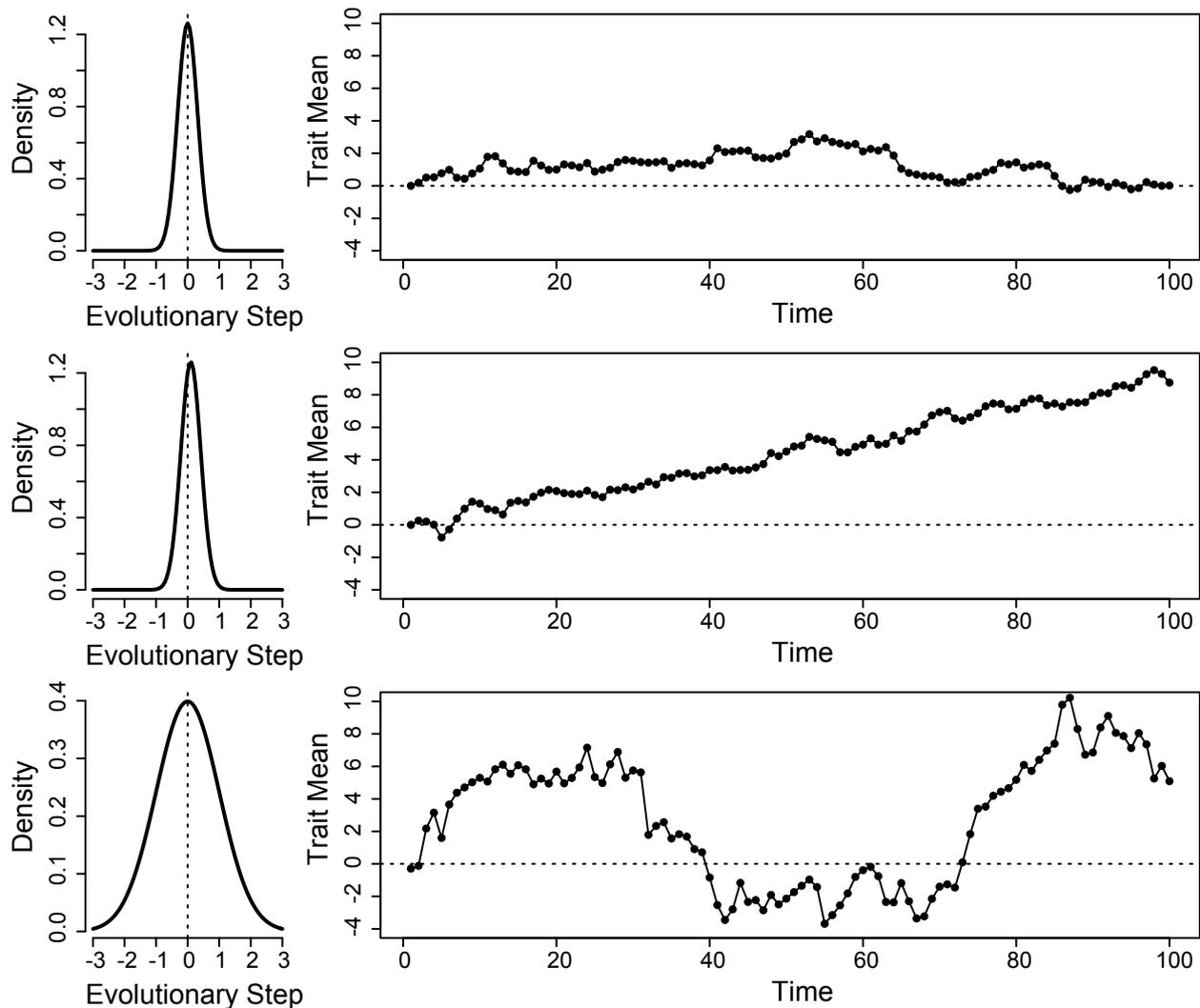


FIGURE 2.—The distribution of evolutionary steps (left column) and realized trait trajectories (right column) for random walks and directional evolution. Top row corresponds to a random walk with a mean step of zero (no trend), and a modest step variance. The middle row indicates a sequence with a mean step that is slightly positive (0.1), which produces a trend of increasing trait values over time. Bottom row illustrates a zero mean step but with a larger step variance than the top row, producing a volatile but not inherently directional trajectory. In the usage of this chapter, the top and bottom rows are random walks, and the middle row shows directional evolution.

We mention this distinction here for terminological reasons: paleontological studies usually talk about random walks whereas phylogenetic ones typically invoke Brownian motion, and it is not widely appreciated that the outcome and parameters from these models are comparable. We'll maintain the convention of applying the term “random walk” to paleontological time-series and “Brownian motion” to phylogenies, but bear in mind that the results are equivalent.

Stasis.—Eldredge and Gould (1972) introduced the term *stasis* to refer to trait trajectories that show little net change over time. Recent depictions of stasis usually show trajectories with modest excursions that quickly return to the long-term mean (Gould, 2002; Eldredge et al., 2005). Sheets and Mitchell (2001) model this notion of stasis as a white-noise process in which evolutionary trajectories have a constant mean, with deviations around that mean that are uncorrelated and normally distributed (Fig. 3). The parameters of this stasis model are the mean and variance of the trait values (Table 1). Note that these differ from the step mean and variance discussed above, which refer to the mean and variance of evolutionary increments (steps).

Evolution in Ancestor-Descendant series and Phylogenies.—All three evolutionary models above—

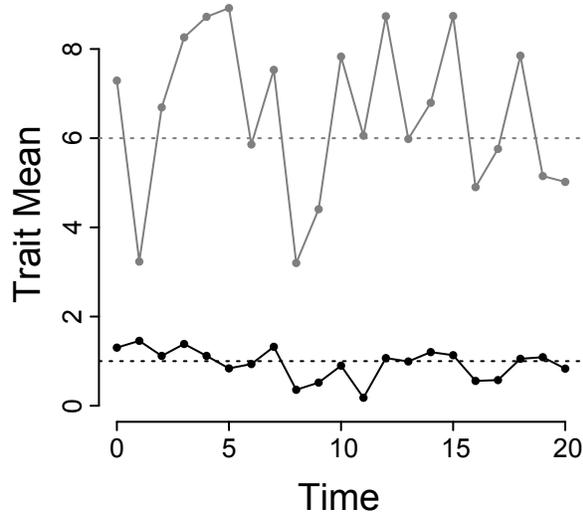


FIGURE 3.—Two trait trajectories generated according to the stasis model, one with little variance around the trait mean (bottom black line), the other with much higher variance (top grey line).

TABLE 1.—Models, parameters and symbols for the three main models in this chapter. Note that mean and variance refer to the actual trait values for the stasis model, but to the distribution of evolutionary increments, or steps, for random walks and directional evolution.

Model	Parameter names	Symbol
Unbiased Random Walk	Step variance	σ_{step}^2
Directional Evolution	Step variance	σ_{step}^2
	Step mean	μ_{step}
Stasis	Trait Mean	θ
	Trait Variance	ω

directional evolution, random walks, and stasis—are described in terms of a single evolving lineage. These notions extend to clades simply by holding that each lineage in the clade shares the same evolutionary dynamic. So, for example, assuming Brownian motion within a clade amounts to assuming that every species within the clade evolves according to this model, with the same parameter value.

WHAT USE ARE MODELS?

Simulation.—One use of models is to simulate different evolutionary outcomes. Given specified values of the evolutionary parameters, one can generate evolutionary trajectories. Using the package *paleoTS*, it is straightforward to simulate ancestor–descendant sequences according to the models discussed in the previous section:

```
Trait1 = sim.GRW(ms=1.0, vs=0.1)
Trait2 = sim.GRW(ms=0, vs=1.0)
Trait3 = sim.Stasis(theta=0,
                    omega=0.1).
```

The function `sim.GRW()` generates directional change if the mean step (argument `ms`) differs from zero and an unbiased random walk if `ms=0`. The step variance is set with the `vs` argument. So, `Trait1` exhibits an increasing trend because the mean step is positive, and `Trait2` is an unbiased random walk. The function `sim.Stasis()` takes two arguments, one indicating the phenotypic position of the mean (`theta`) and the other indicating the variance around

that mean (ω). We did not have to specify the number of samples, their ages, and other conditions because these functions use default values suitable for our purposes. Refer to the help pages for these functions to learn about these defaults and how to change them. The `plot` function can be used to visualize the resulting sequences, for example, `plot(Trait1)`. Analogous functions to simulate trait evolution on phylogenies are available in several packages (e.g., `sim.char()` in the *geiger* package).

Simulations help to guide expectations about which evolutionary outcomes are likely under different conditions. Early on, simulations were very effective in demonstrating how unreliable intuition could be when interpreting evolutionary patterns. This is particularly true for trends, which are very tempting to invoke even when evolution is not truly directional (e.g., Raup and Gould, 1974; Raup, 1977; Bookstein, 1987). If one generates a series of random walks (like `Trait2` above), it will not be long before the human brain finds a sequence that it perceives as directional. Fundamentally, simulations are limited because they are a forward-only approach: given a set of assumptions, they generate data. More often, scientists want to do the reverse—infer something about the processes that generated real data. This is the task of hypothesis testing and model fitting, the subjects of the next two sections.

Null Hypothesis Testing.—The earliest statistical analyses of paleontological sequences used the random walk as a null hypothesis. If observed patterns differed strongly from the expectations of a random walk then this model was rejected, and some other evolutionary explanation was sought. The random walk was used as a null because there was keen interest in documenting directional trends, but it was felt that these trends should be invoked only if the non-directional model of a random walk could be rejected first (Raup, 1977; Bookstein, 1987).

The main shortcoming of these null-hypothesis tests as applied to paleontological series is that they have low statistical power to reject the null hypothesis (high Type II error) (Roopnarine et al., 1999; Sheets and Mitchell, 2001). And, when data are noisy, sampling error can cause incorrect rejection of random walks in favor of stasis (Type I error) (Raup and Crick, 1981; Hannisdal, 2006). Moreover, it is not at all obvious

why the random walk model deserves the privileged null status that allows it to be provisionally accepted unless strongly indicated otherwise (Anderson et al., 2000; Burnham and Anderson, 2010). Stasis, random walk, and directional change are three viable evolutionary models, and it would be better to evaluate them on equal footing. The methods of Gingerich (1993) and Roopnarine (2001) represent attempts in this direction by defining ranges of outcomes that are consistent with each of these models, but they do not allow assessing the performances of different models in statistical terms. To do so requires an explicit procedure for fitting models and comparing how well they account for observations.

Fitting and Comparing Models.—Models have parameters that can take on different values, called adjustable or free parameters. If we want to explain a particular data set, some parameter values will be more consistent with observed features than others. For example, a negative mean step is not compatible with an evolutionary sequence in which a trait increases strongly over time. The process by which optimal parameter values are determined is called *model fitting*. Models may be fit according to different criteria that determine the optimality of fit between model and data.

One very general and powerful approach to model fitting is *maximum-likelihood estimation*. Wang (this volume) discusses the logic of likelihood at length, and we will not go into much detail here. Briefly, likelihood is defined as the probability of obtaining the observed data given a specified model. This approach to model fitting advocates choosing those values for parameters that maximize the likelihood, i.e., that yield the highest probability of producing the data at hand. The natural logarithm of the likelihood, called the *log-likelihood*, is a convenient measure of how well a model fits data.

Models with many parameters usually fit data better than simpler models because complex models have more “knobs” that can be adjusted to better match patterns in the data. Models of differing complexity are said to be *nested* if the simpler model is a special case of the more complex model. For example, a random walk is a special case of directional evolution (as described above) because the latter reduces to the former when the mean evolutionary step is set to zero. For nested models, likelihood can be used in a test that treats the simpler model as the null hypothesis, using a likelihood

ratio test (see Wang, this volume).

A more general approach for comparing models—nested or not—is to use *information metrics* such as the Akaike Information Criterion (AIC). AIC balances goodness-of-fit against model complexity:

$$\text{AIC} = -2\log L + 2K,$$

where $\log L$ is the log-likelihood and K is the number of adjustable model parameters. AIC scores represent the amount of information lost in approximating reality with a model (Akaike, 1974; Anderson et al., 2000; Burnham and Anderson, 2010); because losing information is undesirable, lower AIC scores indicate greater model support. A modified version, AIC_c , performs better when the number of observations, n , per parameter is less than about 40 (Burnham and Anderson, 2010), and it is recommended for routine use:

$$\text{AIC}_c = \text{AIC} + (2K[K + 1]) / (n - K - 1).$$

It is convenient to translate AIC_c scores into Akaike weights by subtracting the score of the best model, $\Delta_i = \text{AIC}_i - \min(\text{AIC})$, and then performing the following calculation:

$$w_i = \exp\left(-\frac{1}{2}\Delta_i\right) / \sum_j \exp\left(-\frac{1}{2}\Delta_j\right).$$

These Akaike weights sum to one across a set of candidate models, and they can be interpreted as the proportional support received by each model. With Akaike weights, model support is measured on a continuum, with no clear distinction between ‘significant’ and ‘non-significant’ support. Because all such divisions are ultimately arbitrary, this property is not necessarily a liability. Rather than focusing on significance cutoffs, Akaike weights and similar metrics encourage scientists to attend to models in proportion to their empirical support.

In order to compute likelihoods, it is necessary to know the probability distribution for possible outcomes. It turns out that, for all the evolutionary models described above, evolutionary changes from ancestor to descendant in a paleontological sequence or along a branch in a phylogeny are expected to be normally distributed, with means and variances that depend on the model parameters and the data. Ancestor–descendant transitions may be analyzed separately, or an entire set

of populations from an evolutionary sequence or phylogeny can be analyzed jointly as a single multivariate normal observation. We do not delve into the actual likelihood functions here; a summary is presented in the online supplementary material and interested readers may consult Hunt (2006; 2008a) for paleontological sequences, and Garland and Ives (2000), Blomberg et al. (2003), and O’Meara et al. (2006) for phylogenetic applications.

ANALYZING ANCESTOR – DESCENDANT SEQUENCES

Data: Size and Shape in Cantius teeth.—In this section, we examine evolution in an ancestor–descendant sequence from the primate lineage *Cantius* reported by Clyde and Gingerich (1994). We focus on two traits of the lower first molar tooth (m1): length and the length-to-width ratio (L/W). These represent a size-related trait and a shape trait, respectively, and both were natural-log-transformed prior to analysis by Clyde and Gingerich.

In R, specialized data types can be stored in objects called *classes*. In the online supplement file, these two data sets are represented as objects of class `paleoTS`, which is essentially a list with vectors containing the means, variances, sample sizes, and ages of all samples, along with some auxiliary information. Entering `cantiusL` at the R command prompt will print these data to the screen. The components of these objects can be accessed just like other lists in R by using the `$` operator. For instance, the vector of means for m1 length is `cantiusL$mm`. The help page for `as.paleoTS()` documents the `paleoTS` class; see also the online supplement to this chapter.

These two time series show very different evolutionary patterns (Fig. 4). The length of m1 shows a large net increase over time that gives the strong visual impression of a trend (Fig. 4.1). The magnitude of net increase is 0.3 log mm. Given that the average within-sample standard deviation is 0.038 log mm, this increase represents a size gain of nearly 8 standard deviations, which is substantial. In contrast, the shape trait, L/W of m1, shows limited variation around a more-or-less stable mean (Fig. 4.2). Qualitatively at least, size seems to evolve directionally, whereas shape conforms to the stasis model.

Null Hypothesis Testing.—Paleontologists have used various null hypothesis tests that key on different aspects of the behavior of random walks. To illustrate the null-hypothesis testing approach, in this section we implement two tests described by Raup and Crick (1981). The first test relies on tallies of the number of trait increases and decreases over an evolutionary sequence. These counts are compared to the binomial probability of obtaining the observed number of increases and decreases, assuming both outcomes are equiprobable. L/W of m1 shows 19 increases and 20 decreases, and a binomial test, `binom.test(x=19, n=39)`, confirms that there is no systematic tendency for this trait to increase or decrease ($P = 1.0$). The same conclusion is reached for the seemingly trended character, m1 length. This time-series has 19 increases and 15 decreases, an outcome that is not particularly surprising when increases and decreases are equally probable ($P = 0.61$). The five intervals in which size showed no change at all (at the resolution of measurement precision) were omitted for this test as per Raup and Crick (1981).

The above test considers only the direction of evolutionary changes from one sample to the next, and thus ignores the magnitudes of change. This potential shortcoming can be addressed by testing if the average evolutionary change across all increments differs from zero. In a strongly trended sequence, the mean increment should differ significantly from zero. The difference between adjacent means can be computed using the `diff()` function, and a t-test can evaluate these differences. Applying this approach to the m1 L/W, `t.test(diff(cantiusLW$mm))`, yields a mean evolutionary difference that is very close to zero ($P = 0.97$). The mean evolutionary increment for m1 length is positive (0.0077 log mm) but again not large enough to reject the null hypothesis of no trend ($P = 0.24$). (Note when the `t.test()` function is given a single vector as an argument, by default it performs a one-sample test with the null hypothesis that the mean of that sample is equal to zero.)

These two tests find no strong evidence for directional evolution in tooth size or shape. This result was expected for m1 L/W, but may be surprising for m1 length because Figure 4.1 really does look like a trend. There are two possible explanations for the apparent lack of directionality in m1 size: either our intuition of directionality is faulty, or the tests employed have

little power to reject the null hypothesis of a random walk, even when it is false. Next, we'll use AIC_c scores to evaluate support for models directly to distinguish these possibilities.

Model fitting.—Via maximum-likelihood, we can fit the three models of stasis, random walk, and directional change to these two trait trajectories. There are two available parameterizations for these models, depending on whether one looks at ancestor–descendant differences or the joint distribution of trait values for all samples (Hunt, 2008a, see also the online supplement to this chapter). The latter approach is better able to correctly identify trends in noisy data (Hunt, 2008a). Accordingly, we will focus on this approach, although the ancestor–descendant parameterization leads to

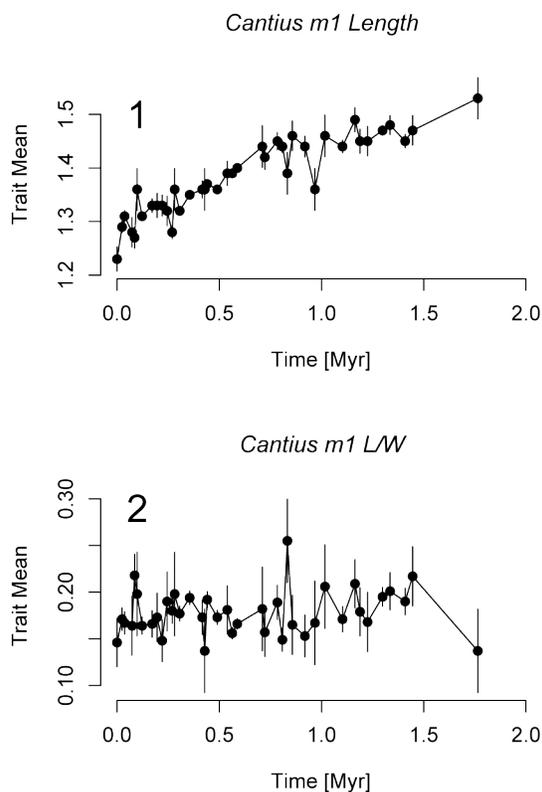


FIGURE 4.—Evolutionary trajectories in dental measurements for the *Cantius* lineage. 1, length of the lower first molar. 2, length to width ratio of this same tooth. Time is in millions of years from the start of the sequence and error bars represent one standard error on the mean.

conclusions that are qualitatively similar. The three models can be fit and compared using the appropriate function from the *paleoTS* package:

```
fit3models.joint(cantiusL)
fit3models.joint(cantiusLW) .
```

The outputs from these two function calls are summarized in Table 2. For m1 length, directional evolution merits substantially more model support than the random walk model (higher log-likelihood, lower AIC_c). The directional evolution model outperforms the other models so well that it receives 92% of the available evidential support. As expected, the stasis model fits the data very poorly and receives negligible support. Nearly the opposite pattern obtains for m1 L/W in that stasis is the best-supported model, although the random walk also receives moderate support (Akaike weight = 0.33). The rather strong model support for directional evolution in m1 length suggests that the non-significant results from the null-hypothesis tests reflect their low statistical power.

In addition to gauging the relative support for different models, parameter estimates can inform us about evolutionary dynamics. The maximum-likelihood estimate of the mean step is 0.13 log mm per Myr (units are per Myr because time was recorded in millions of years); this is a measure of the strength of the evolutionary trend (see Fig. 4.1). The step variance is high enough so that there is considerable variation in the evolutionary increments, thus generating a trend that is consistent but not uniform. Sampling error—noise in estimating sample means—will also cause point-to-point fluctuations in fossil sequences.

Regardless of the mode of evolutionary change, the step variance parameter of the random walk model is a useful measure of evolutionary rate because larger step variances produce greater evolutionary divergences (Hunt, 2008a). However, this parameter is only comparable for traits measured in the same units of morphology and time. To allow for comparisons across different traits, we can divide the estimated step variance by the within-sample variance, pooled across all samples in a sequence. This scaling is analogous to measuring traditional rates using standard deviation units (i.e., in haldanes; see Gingerich, 1993), and the resulting quantity is proportional to the rate metric proposed by Lynch (1990) (see Geary et

al., 2010). Because the step variance and population variance are in squared units, we can take the square root to convert the rate metric back to the dimensionality of the original data (Gingerich, 2009). Applying these calculations to the values reported in Table 2 yields a rate parameter that is over three times higher for m1 length (1.92 sd per Myr) than for m1 L/W (0.58 sd per Myr), consistent with the much greater evolutionary changes observed in the former trait.

ANALYZING POPULATIONS RELATED BY PHYLOGENY

Data: Femoral Dimensions in Dinosaur Clades.—In this section we demonstrate a variety of analyses of femoral dimensions in several dinosaur clades. The femoral data are from Carrano (2006), updated with more recently available data from the literature. For most analyses, we focus on the natural logarithm of femur length, but we add the logarithm of mediolateral femoral diameter as a second characteristic when demonstrating methods that examine relationships between traits. As in Carrano (2006), the phylogenetic hypothesis for the Dinosauria is a composite from several smaller studies, again updated to reflect more recent phylogenetic work on the group.

In addition to topology, models of phenotypic evolution require information about branch lengths, analogous to the requirement of an age model for ancestor-descendant series. Biologists often measure branch lengths in units of molecular distance produced by phylogenetic inference of DNA sequences, sometimes calibrated to geologic time (Sanderson, 2003; Drummond et al., 2006). Although molecular branch lengths are unavailable to paleontologists, stratigraphic information can provide age constraints because each node must be at least as old as the oldest terminal taxon it includes (assuming the age assignments and topology are correct, of course). Without further modification, however, this procedure will result in a tree with branches of zero length between nodes and their oldest included terminal (Brusatte et al., 2008). Because most models of evolution predict no change over branches of zero length, this would amount to assuming that the oldest sampled taxon in a clade represents the exact morphology of the ancestral state of that clade. Although paleontologists can expect to sample ancestors sometimes (Foote, 1996), in the absence of

TABLE 2.—Model-fitting results for the *Cantius* sequences. Akaike weights for models receiving at least moderate model support (> 0.2) indicated in bold. The phenotypic variances pooled across samples are 0.0015 (m1 length) and 0.0020 (m1 L/W).

Trait	Model	logL	AIC _C	Akaike weight	Parameter estimates
m1 length	Directional	88.76	-170.84	0.92	$\mu_{step} = 0.13, \sigma_{step}^2 = 0.003$
	Random Walk	85.11	-165.89	0.08	$\sigma_{step}^2 = 0.0055$
	Stasis	49.24	-94.16	0.00	$\theta = 1.38, \omega = 0.0045$
m1 L/W	Directional	96.60	-186.55	0.12	$\mu_{step} = 0.012, \sigma_{step}^2 = 0.0007$
	Random Walk	96.47	-188.61	0.33	$\sigma_{step}^2 = 0.0007$
	Stasis	96.99	-189.67	0.56	$\theta = 0.18, \omega = 9.4 \times 10^{-5}$

character data it would be imprudent to assume this is always the case (Fisher, 1994; Smith, 1994; Marcot and Fox, 2008).

There are several possible ways of adjusting these zero-length branches, including setting an arbitrary minimum length, adding length proportional to cladistic character change (Ruta et al., 2006), or possibly from information about preservation rate for the clade. For all the examples here, we follow the approach of Brusatte et al. (2008), in simply pulling the root back in time by a fixed amount, and distributing the added time across internal branches. We used the `date.phylo()` function written by Graeme Lloyd (available at <http://www.graemetlloyd.com/>) to set the origin of the Dinosauria to be 20 Myrs before the earliest dinosaur fossil. This range extension resulted in reasonable times of origin for the major dinosaurian clades. The ages of the terminal taxa were set as the midpoint of the stratigraphic stage or stages in which the taxon existed, using the ages for stage boundaries in the GSA 2009 time scale (Walker and Geissman, 2009). Handling the zero-length branches that result from stratigraphically calibrated trees is a surprisingly tricky problem, and we refer readers to the discussion in Brusatte et al. (2008, online supplement) for a review.

Fitting Models: Cope's Rule in Ornithischians.—The enormous sizes attained by many dinosaur taxa, coupled with the small size of most early dinosaurs, have long led to suggestions of a systematic trend of increasing body size (Cope's Rule) for this clade. Carrano (2006) tested this notion by analyzing log-femur length in the major clades of dinosaurs. Log-femur

length is a useful proxy for body mass because it is a load-bearing structure in all dinosaur groups, and is known in extant terrestrial vertebrates to exhibit a linear relationship with the log of body mass. Carrano's analyses found different patterns in different clades, but evidence for overall increases in size in many groups of dinosaurs. This general conclusion was also supported by Hone et al. (2005), who made phylogenetically informed comparisons of older and younger taxa and found that the younger taxon was usually the larger of the two. The results of both of these studies support Cope's Rule in the broad sense of implying directional trends toward larger body sizes.

Although both of these analyses incorporate phylogenetic information, they do so in a limited way by focusing on the behavior of whole clades, or select sister-taxon comparisons. If Cope's Rule holds, however, one can simply fit using likelihood the model of directional change directly, using body size values and a phylogeny with branch lengths (Pagel, 2002, see also the online supplement). The signature of directionality is trait values that show a systematic trend with elapsed time from the root of the phylogeny. This calculation does not require estimating the ancestral states of nodes, which is convenient because this can only be done with considerable uncertainty (see below).

We demonstrate this approach by comparing two models: (i) Brownian motion, and (ii) Brownian motion with a trend; this is equivalent to the comparison between a random walk and directional evolution in the *Cantius* analyses. If Cope's Rule is pervasive, then the trend model should receive much more model support than Brownian motion, and its directionality parameter

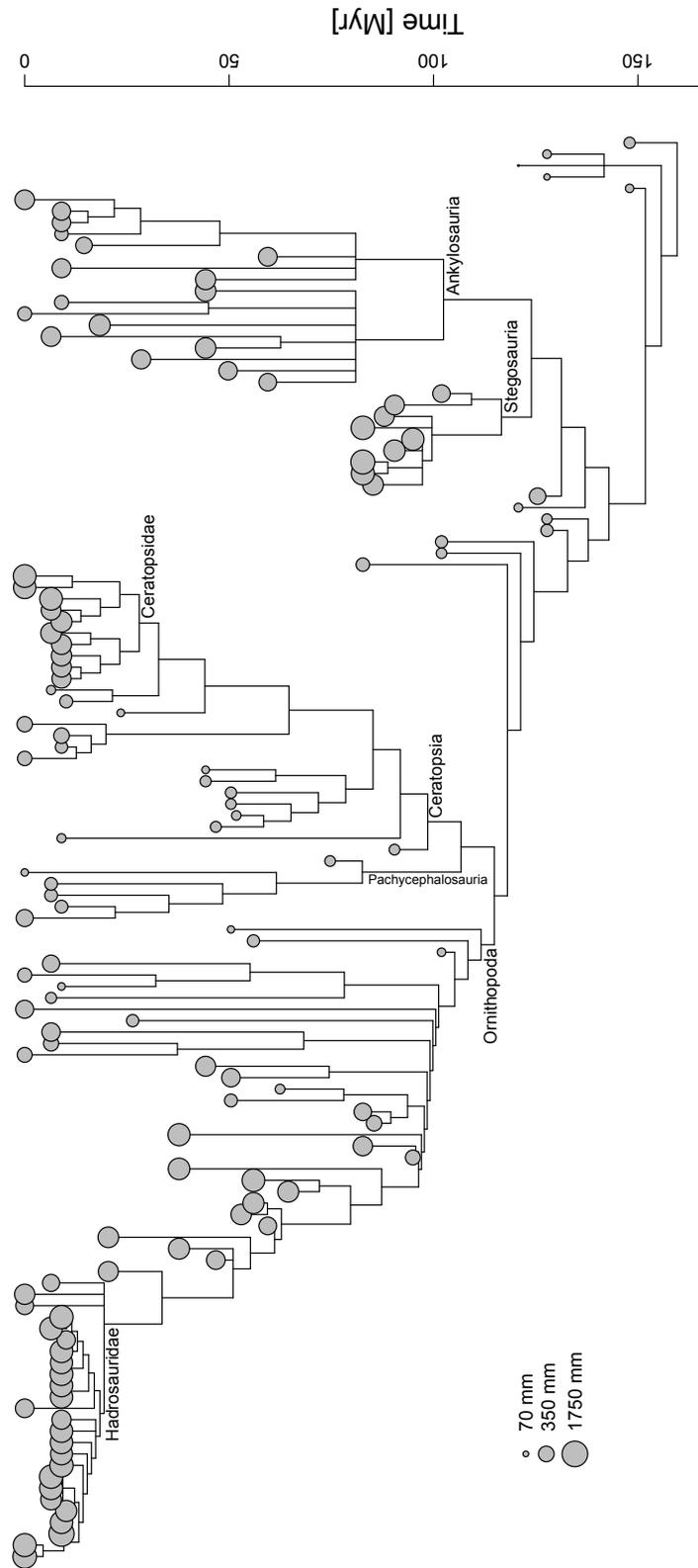


FIGURE 5.—Phylogeny of ornithischian dinosaurs from the composite tree of Carrano (2006) modified according to other studies (Ford and Kirkland, 2001; Weishampel et al., 2003; Novas et al., 2004; Vickaryous et al., 2004; Averianov et al., 2006; Ryan, 2007; Carpenter et al., 2008; Maidment et al., 2008; You et al., 2008; Arbour et al., 2009; Boyd et al., 2009; Dalla Vecchia, 2009; Sues and Averianov, 2009; Butler et al., 2010), with branch lengths scaled to geological time. Taxon names are omitted for clarity, although a few of the larger named clades are labeled below the appropriate node. Symbol sizes are scaled to log femur length. Time axis extends backwards from the youngest terminal taxa, in millions of years.

should be positive, indicating increasing sizes over time. We apply this approach to one large clade of dinosaurs, the Ornithischia (Fig. 5).

We can use the `fitContinuous()` function from the *geiger* package to fit these two models. Assuming the phylo object is called `OrnithTree` and log-transformed femora lengths are called `FemL`, the relevant R commands are:

```
fitContinuous(OrnithTree, FemL,
              model='BM', meserr=0.014)
fitContinuous(OrnithTree, FemL,
              model='trend', meserr=0.014).
```

The last argument to this function, `meserr`, requires some explanation. This value represents the *standard error* of the measured trait (`FemL`), which is relevant because sampling error in estimating `FemL` contributes to the observed differences between species. Without accounting for this contribution, we could mistake noise for true differences and thus overestimate evolutionary divergence. Here, each species is represented by a single femoral measurement that we can consider as an estimate of mean adult femoral length. The standard error of a species mean is equal to the standard deviation of trait values in that species, divided by the square root of the number of observations, here set to be one. (This is an approximation here because a few taxa are associated with several femora and we analyzed only the largest of these. For the sake of simplicity, we ignore this complication.) Four dinosaur species for which more than five adult-sized femora are known (*Iguanodon bernissartensis*, *Dryosaurus lettowvorbecki*, *Tyrannosaurus rex*, *Allosaurus fragilis*) have standard deviations of log-femur length that range from 0.011 to 0.019 log mm. We set the standard error of all species to the average of the four species' values (0.0144 log mm). This approach is necessarily rough, but it is almost certainly better than assuming that sampling error is absent. Note that sampling error captures the uncertainty in estimating a quantity from a finite sample, and therefore it differs from the measurement error that arises from imprecision in capturing data from a specimen (e.g., from slight differences in caliper placement). Unless sample sizes are very high, measurement error is usually much smaller than sampling error, and so the latter is of greater concern. Although it is invisible to the user, the `fit3models`

`joint()` function applied above to the *Cantius* data similarly accounts for sampling error.

The model support and parameter estimates for these two models are summarized in Table 3. The trend model has a higher log-likelihood and a lower AIC_c score than Brownian motion, and it accounts for most of the Akaike weight. The mean step is positive and implies an average increase in log femur length of 0.0063 log mm per Myr, i.e., $\log(\text{femL}_1) - \log(\text{femL}_0)$ is 0.0063, where the subscripts indicate time in millions of years. This is equivalent to $\log(\text{femL}_1/\text{femL}_0) = 0.0063$, and taking the exponential of both sides shows that the expected ratio of femur lengths before and after one million years is $\exp(0.0063) = 1.0063$, which translates to about a 0.6% increase per Myr. The dominance in Akaike weight of the trend model and the positive mean step are consistent with Cope's Rule in this clade. Over the approximately 160 Myr history of the Ornithischia, the predicted increase in femur length, $\log(\text{femL}_{160}) - \log(\text{femL}_0)$, is $(160)(0.0063) = 1.008$ log mm, equivalent to a nearly three-fold difference ($\exp[1.008] = 2.74$).

The trend model can be summarized on a plot of log-femur length versus the time elapsed from the root of the clade to each terminal taxon (Figure 6). Each dot is a species, and the dotted line shows the result of the fitted trend. The slope of this line is equal to the mean step (0.0063 log mm per Myr), and the intercept is equal to the estimated femur length at the root of the phylogeny. This uniform trend can be compared to a locally weighted (lowess) regression (Cleveland, 1979) through these data. Lowess regressions offer a convenient way to run a curve through a scatter plot without committing to any particular functional form. This curve, indicated by the solid line in Figure 6, also increases over time, consistent with a Cope's Rule pattern. However, the lowess curve is initially quite steep but then nearly levels off. This pattern suggests it might be fruitful to consider additional models in which the strength of Cope's Rule diminishes over time.

The Brownian motion and trend models are comparable to the random walk and directional evolution models, respectively. As a result, we can compare the parameter estimates between these sets of models, as long as the traits in question are in equivalent units. The strength of the body-size trend in ornithischians, 0.0063 log mm per Myr, can be compared to that in *Cantius*, 0.13 log mm per Myr. The directionality is stronger

TABLE 3.—Model-fitting results for Ornithischia log femoral lengths comparing Brownian motion (BM), BM with a trend and stasis (white noise).

Model	logL	AIC _c	Akaike weight	Parameter estimates
BM with trend	-86.60	179.39	0.87	$\mu_{step} = 0.0063, \sigma^2_{step} = 0.0098$
BM	-89.59	183.27	0.13	$\sigma^2_{step} = 0.0103$
Stasis (white noise)	-153.33	310.75	0.00	$\theta = 6.07, \omega = 0.68$

by more than a factor of 20 in the *Cantius* data, but we must consider the temporal durations of the two case studies. Ornithischians existed for over 160 million years—roughly 100 times longer than the *Cantius* time series—and it is unlikely that very strong trends will persist for long stretches of geological time. Starting from the estimated ancestral ornithischian femur size and applying a trend as strong as *Cantius* over 160 Myrs results in an average femur length that exceeds 127,000 km! As Gingerich (1983; 1993; 2009) has emphasized, we need to be very careful about comparing patterns and rates over very different durations. It is safer to compare the parameter estimates among clades of similar duration.

One particularly useful parameter to compare is the step variance of the Brownian motion model. Because this parameter determines the magnitude of evolutionary change under BM, it is a useful measure of the rate of phenotypic evolution. In fact, in phylogenetic studies, this parameter is the standard metric for measuring evolutionary rates (O'Meara et al., 2006). If we compare the step variance for ornithischian femora to that from another dinosaur clade of comparable duration and diversity, the Sauropodomorpha, we find that the rates of change are almost identical across the two clades (0.0103 log mm² per Myr for Ornithischia, 0.0095 log mm² per Myr for Sauropodomorpha). Although the detailed patterns of size evolution differ between the two groups (Carrano, 2006), both clades exhibit similar overall rates of body size evolution.

The last line in Table 3 gives the statistics for a model of stasis (white noise). The term stasis is usually applied to individual lineages rather than entire clades, and so its use here requires some comment. For a clade, this model implies that species attributes can be thought of as random, independent draws from a normal distribution with a stable mean and variance. The key aspect is not the magnitude of divergence, but rather the independence of species—closely related taxa are no more similar to each other than distantly related

ones, and different subclades show no systematic trait differences. Both of these properties do not hold for ornithischians: there are systematic differences in log femur length across clades (Fig. 5) and time intervals (Fig. 6), and closely related taxa are often similar in femoral dimensions (Fig. 5). As a result, this model fares quite poorly – its log-likelihood is low and its AIC_c is very high (Table 3). Biologically, we might expect this kind of pattern when traits evolve in a bounded space and evolutionary steps are large relative to the range of available morphologies (see Sidlauskas, 2008 for an empirical example of this kind of pattern). Although stasis is seldom modeled explicitly at the clade level, this model forms an end-member for several other approaches (see the Other Models and Other Issues sections).

Relationships Among Traits.—One common goal in paleontology and biology is to assess the relationship between two different attributes, such as brain size and body size. In this section, we will consider femoral length and mediolateral diameter as our two traits of interest, focusing on a clade of ceratopsian dinosaurs (Fig. 7). We wish to explore the relationship between these two traits, accounting for the fact that our species are related by genealogy.

A plot of log-femoral length versus log-mediolateral diameter shows that the two are positively correlated across species (Fig. 8). A least-squares regression through these data (Fig. 8, solid line) is significant ($P = 0.0005$). However, the phylogenetic dependence among species violates the least-squares assumption that residuals are independent. The two main solutions to this problem take different strategies for handling the phylogenetic dependence among species. The method of independent contrasts creates a new set of variables that are independent, whereas generalized least-squares accounts statistically for the non-independence among terminal taxa. The following paragraphs will describe and apply each of these approaches to the data in Figure

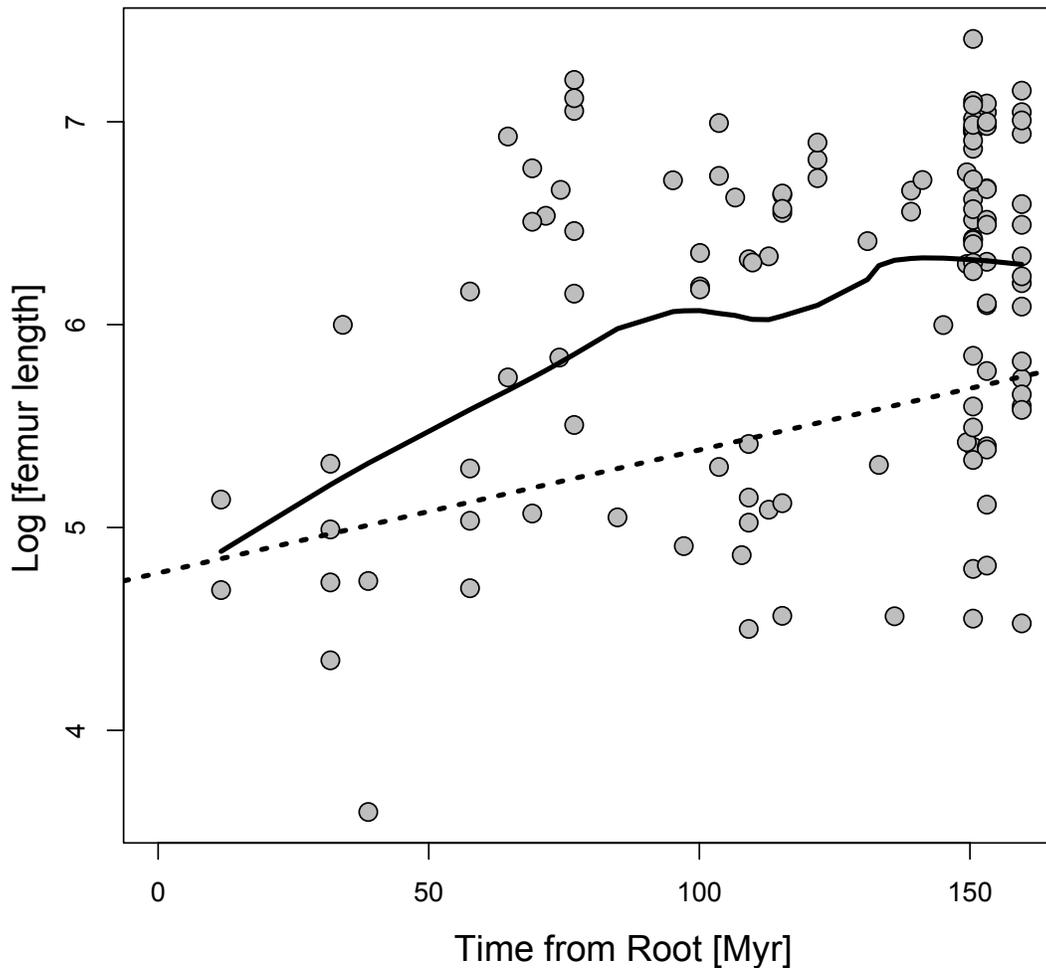


FIGURE 6.—Femur length for each ornithischian species plotted as a function of the elapsed time from the root of the tree. Dotted line shows the trend in size implied by the directional (BM with trend) model. Solid line is a lowest regression through the data points.

7. The R code to perform these analyses is set off in Box 1.

Felsenstein (1985) developed *phylogenetically independent contrasts* (PIC) as the first truly tree-based approach for analyzing relationships among traits. This procedure starts with a pair of sister taxa, say *Chasmosaurus belli* and *Agujaceratops mariscalensis*, and calculates the difference between their femoral lengths (see Fig. 7). This difference is called a *contrast*, and it does not matter which value get subtracted from which, so long as the order is consistent from trait to trait. Then, the trait value for the corresponding node (node 16) is estimated. Without getting into details (Felsenstein, 1985; 2004), node values are estimated assuming Brownian motion and amount to a weighted

average of all terminal taxa tipward of the focal node. Next, a contrast is created between this node value and the next unit out on the tree (here, *Pentaceratops sternbergii*). The process of estimating states for nodes and computing contrasts is repeated until all taxa are included. A fully bifurcating tree with N terminal taxa yields $N - 1$ contrasts, one for each internal node in the phylogeny. Each branch of the tree is included in exactly one contrast, and therefore the contrasts are all phylogenetically independent of each other. Moreover, by dividing each contrast by its expected standard deviation (Felsenstein, 1985; 2004), we get a set of independent, standard normal variates (mean of zero, unit variance), appropriate for any statistical analysis. The whole process is repeated for the second

trait, mediolateral diameter, and the paired contrasts can be plotted (Fig. 9).

The regression between these contrasts is positive and significant ($P=0.004$), implying that evolutionary changes in femoral length are positively correlated with evolutionary changes in mediolateral diameter. Because the direction of each contrast is arbitrary, this regression should be constrained to go through the origin (Garland et al., 1992) (Box 1). Both femur dimensions increase or decrease in concert across contrasts except for the one between the narrow femur of *Centrosaurus apertus* and the more typical femur of its sister taxon, *Monoclonius crassus* (Fig. 9, node 13; see Fig. 7).

The second method for dealing with phylogenetic dependence is *generalized least squares* (GLS). GLS

does not attempt to create a new set of independent variables, but instead seeks to correctly account for their non-independence. This is done by specifying the covariance among data points, which GLS uses to downweight the joint influence of points that have high covariance (and thus are partially redundant). The predicted covariance among taxa comes from the phylogenetic tree coupled with a model of phenotypic evolution. Brownian motion is almost universally assumed, but other models may be implemented as well (Hansen and Martins, 1996; Martins and Hansen, 1997). Under BM, the expected covariance among terminal taxa in a phylogeny is proportional to their shared path length from the root of the tree. Taxa that share most of their path from the root, such as *Achelousaurus horneri* and *Pachyrhinosaurus lakustai* (Fig. 7), have high covari-

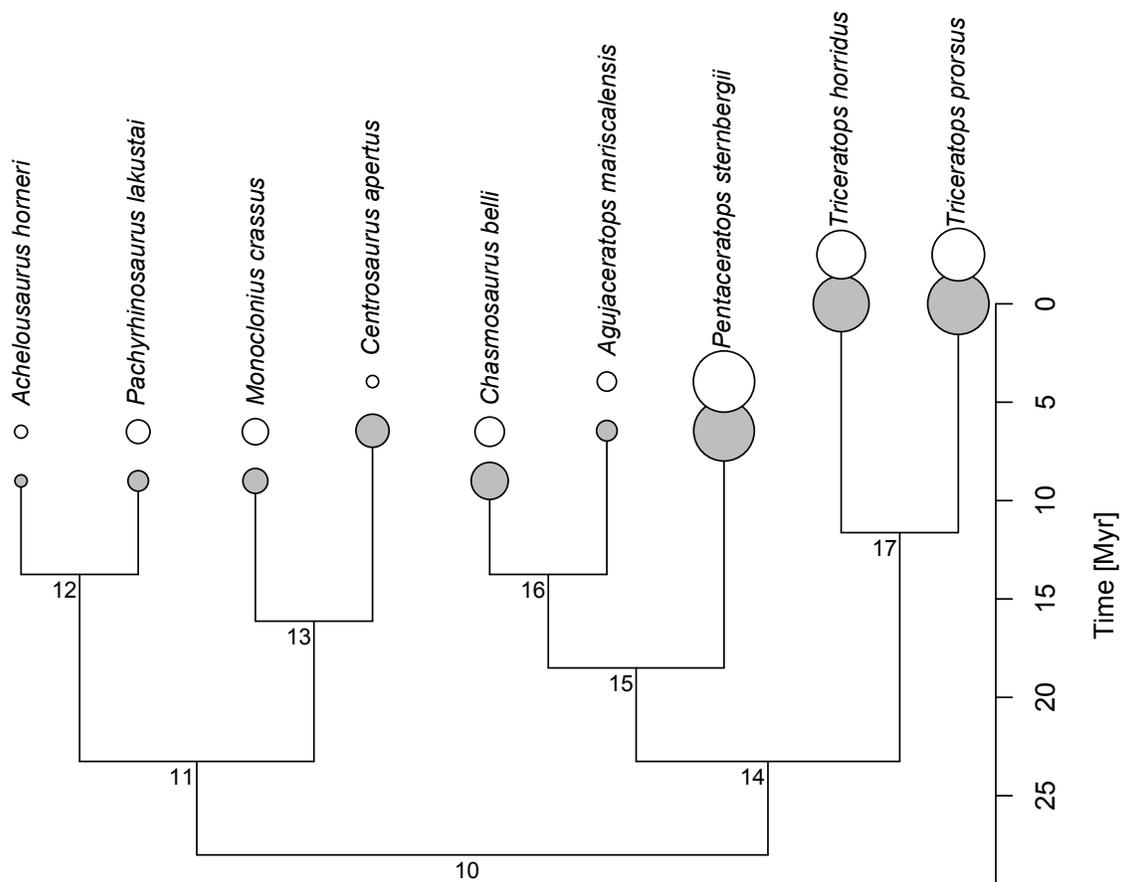


FIGURE 7.—Phylogeny of a clade of ceratopsian dinosaurs. Grey circles scale with log-femur length and white circles scale with mediolateral diameter of the femur. Some species were excluded because they lacked data for one or both of these features. Axis on the right indicates time, in millions of years, counting backwards from the youngest terminal taxa. Internal nodes are labeled by number.

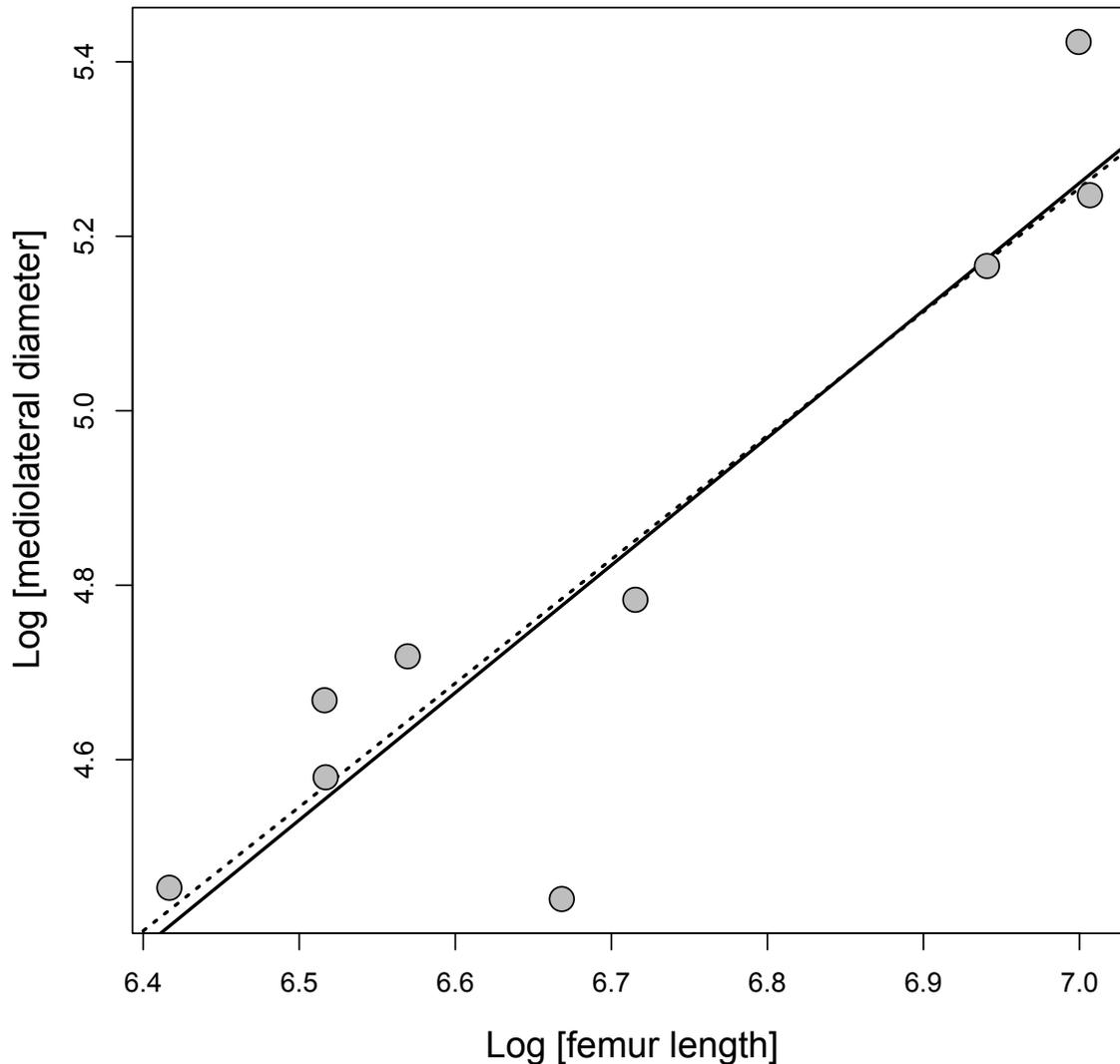


FIGURE 8.—Bivariate plot of mediolateral diameter versus length of the femur. Each symbol is a species from Figure 7. Solid line is the least-squares regression, which is inappropriate for these data because species values are not independent. Dotted line shows the GLS regression, which takes into account the dependence due to phylogeny.

ance relative to more distantly related taxa, such as *A. horneri* and *Tricerotops horridus* (Fig. 7); see the online supplement for an example of a phylogenetic covariance matrix.

The slope and intercept of the GLS regression are rather similar to the least-squares results (Fig. 8, dotted line), but the standard errors are larger and the P -value is correspondingly higher, though still significant ($P = 0.0043$). The P -value and slope of the GLS regression are the same as those from the PIC analy-

sis, which should be the case because these methods have similar underpinnings (Garland and Ives, 2000; Rohlf, 2001). That the PIC and GLS analyses are more conservative than ordinary least-squares is reasonable because the phylogenetically informed approaches correctly recognize that non-independence translates into effective sample sizes that are lower than the number of terminal taxa.

GLS is a very flexible approach, able to assess all kinds of linear relationships among variables. For

BOX 1—PHYLOGENETIC INDEPENDENT CONTRASTS (PIC) AND GENERALIZED LEAST SQUARES (GLS) IN R

In the code that follows, we assume the following R objects exist: a phylo object, `CeraTree1`, of the relevant phylogeny, log-transformed femoral length (`FemL.cerat1`) and mediolateral diameter (`MLD.cerat1`), all of which are available from the R data file in the online supplement. In R, any text following the pound sign (#) is interpreted as a comment, and is thus ignored.

```
## Least-squares regression for comparison ##
reg.ols = lm(MLD.cerat1 ~ FemL.cerat1) # lm() fits linear models
print(summary(reg.ols))

## now, do a GLS regression ##
library(nlme) # nlme package has the gls() function
library(ape) # requires ape package
VV = vcv(CeraTree1) # phylogenetic var/covariance matrix
tipheights = diag(VV) # diagonals are distance from root
reg.gls = gls(MLD.cerat1 ~ FemL.cerat1, correlation=corBrownian(
  phy=CeraTree1), weights=varFixed(~tipheights))
print(summary(reg.gls))

## do phylogenetic independent comparisons (PIC) ##
pic.Fem = pic(FemL.cerat1, CeraTree1) # create contrasts for FemL
pic.MLD = pic(MLD.cerat1, CeraTree1) # create contrasts for MLD
reg.pic = lm(pic.MLD ~ pic.Fem - 1) # regress through origin
print(summary(reg.pic))
```

In R, models are specified with the ‘~’ operand, so the code `y ~ x` should be read as “y is a linear function of x.” Including and excluding terms is done with ‘+’ and ‘-’, respectively. For example, `y ~ x1 + x2` reads as “y is a linear function of x1 and x2;” if x1 and x2 are continuous variables, this model corresponds to a multiple regression. In the last line of the PIC code above, “- 1” excludes the intercept from the model, forcing the regression to pass through the origin.

For the GLS regression, it is necessary to specify a correlation structure, and a vector of variances. The function `corBrownian()` computes the expected correlation structure, assuming a Brownian motion model of evolution. Under this model, the variance of evolutionary outcomes is related to duration of time between the root and each terminal taxon, which we computed as `tipheights`. If all taxa were of the same age (e.g., all extant), the `weights` argument need not be included.

example, we could include additional independent variables to the model, and fit the resulting multivariate regression. These predictor variables can be continuous or discrete factors, or any mixture of the two. Note that the dependent variable must be continuous because the residuals are assumed to be Gaussian; violations of this assumption can be handled by general estimating equations (Paradis and Claude, 2002). Finally, we note that GLS is a general statistical approach for dealing with any kind of non-independence, not only

that resulting from phylogeny. For example, there are GLS approaches for accounting for spatial or temporal autocorrelation (Diniz-Filho et al., 2003; Hunt et al., 2005). Sometimes, the term phylogenetic generalized least squares (PGLS) is used to differentiate this approach from GLS used in other contexts.

Ancestral State Reconstruction.—Early in the history of comparative methods, much attention was focused on reconstructing the phenotypes of unsampled

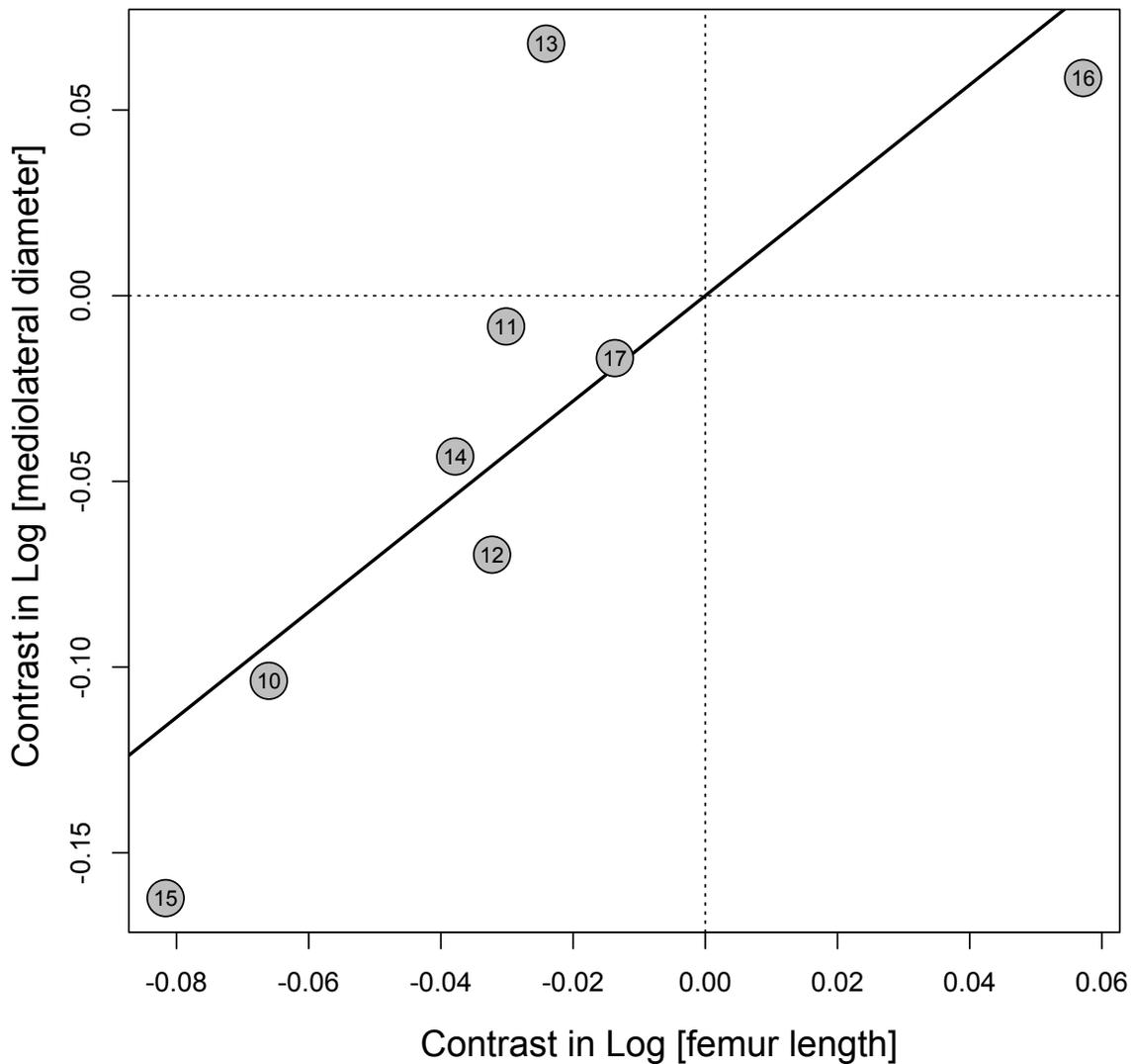


FIGURE 9.—Bivariate plot of the independent contrasts of mediolateral diameter to length of the femur. Each symbol represents a contrast, labeled according to its relevant node in Figure 7. Solid line is the least-squares regression line, constrained to go through the origin. Dotted lines indicate zero values for each contrast.

ancestors. For continuous traits, parsimony-based methods were developed that computed ancestral states that minimized the sum of absolute changes (linear parsimony) or squared changes (squared-change parsimony) across the tree. Although squared-change parsimony can be linked to Brownian motion (Maddison, 1991), parsimony methods are essentially heuristic. Neither is explicitly implemented in R, although both can be performed using the software Mesquite (Maddison and Maddison, 2009).

The two main model-based approaches for reconstructing ancestral states are maximum-likelihood (ML, Schluter et al., 1997) and phylogenetic independent contrasts. GLS can also produce estimates for ancestors (Martins and Hansen, 1997) but these will closely correspond to ML estimates when Brownian motion is assumed (Rohlf, 2001). PIC ancestral states were described above—they are a weighted average of all terminal taxa tipward from a focal node. ML ancestral states differ in that the reconstructed state for a node

is informed by all the terminal taxa, not just the subset of terminals that descend from that focal node. PIC ancestral states are mostly a by-product of analyzing contrasts, and are not commonly used when the goal is simply to estimate ancestral states. Accordingly, we'll focus the rest of this section on ML estimation of unsampled ancestors.

The function `ace()` from the *ape* package can be used to estimate ML ancestral states:

```
ace(FemL.cerat2, CeraTree2,
    method='ML').
```

Sizes of the circles in Figure 10 increase linearly with log-femur length of the terminal taxa, and circles at internal nodes represent ancestral femoral sizes as inferred by ML. The reconstructed value for the root (most basal node) is 5.42 log mm, which is smaller than the mean of all species in this clade, 6.29 log mm. This occurs because the influence of terminal taxa decreases with their distance from a node. In this clade, stratigraphically old species such as *Graciliceratops mongoliensis* have relatively smaller femora, and their disproportionate influence lowers the estimated femur length at the root. In the limit, taxa barely separated from the root have had little time to accumulate evolutionary differences, and they therefore should be rather similar to the ancestral state. A corollary to this reasoning suggests that estimates of the ancestral states for extant species can be greatly improved with the inclusion of even a small number of fossil taxa, especially if they are stratigraphically old (Finarelli and Flynn, 2006).

What is the best approach for estimating ancestral states? Some studies tackle this problem empirically, by comparing inferred and fossil values (Webster and Purvis, 2002a; 2002b; Finarelli and Flynn, 2006), but these are too few at present to offer general conclusions. More broadly, one should consider whether ancestral states need to be estimated at all. Although they have heuristic value, it is important to remember that ancestral states are not observations, but are instead inferences subject to broad uncertainty (Schluter et al., 1997). The degree of uncertainty can be illustrated by examining the confidence intervals output from the `ace` function above. For example, the 95% confidence interval for the ancestral log femur length ranges from 4.99 to 5.83 (148 – 341 mm on the original scale),

roughly a third of the range of log femoral lengths among the terminal taxa. Methods such as model fitting and GLS address interesting biological problems without relying on estimates for ancestral nodes. The use of PIC does rely on ancestral state estimates, but the uncertainty in the estimates is correctly incorporated into the standard estimates of the contrasts (Felsenstein, 1985). Thus, unless the question of interest specifically focuses on the phenotype of ancestors, it is often better to employ methods that do not explicitly rely on their estimation.

OTHER MODELS

In this chapter, we have focused on models that are commonly used and straightforward to understand. Without treating them in detail, it is useful to mention a few of the important models that we have not covered.

Elaborations for Heterogeneous Dynamics.—The models described in this chapter thus far are homogeneous in that they apply uniformly to all populations in a lineage or clade. A variety of approaches have been developed to introduce biologically interesting heterogeneity into evolutionary dynamics. For ancestor–descendant sequences, modest extensions of the methods described here can model punctuations and other kinds of shifts in evolutionary dynamics within a sequence (Hunt, 2008b). For species within clades, many of the heterogeneous models are elaborations of Brownian motion. These include modifications to allow shifts in the BM rate parameter along specified branches (O'Meara et al., 2006), or temporal trends of accelerating or decelerating rates of change (Blomberg et al., 2003). One can also consider models with instantaneous shifts to new means along specific branches, with constant or varying rate parameters across the tree (Thomas et al., 2009). Other methods allow for the evolutionary relationship among traits to change at different points in the tree (Revell and Harmon, 2008; Revell and Collar, 2009).

Ornstein-Uhlenbeck (OU) Models.—Ornstein-Uhlenbeck (OU) models contain a phenotypic optimum that acts as an evolutionary attractor. Populations evolve very rapidly towards the attractor initially, but then decelerate, producing an exponential convergence to the optimum. Populations already located near the phenotypic optimum will exhibit stasis as the attractor

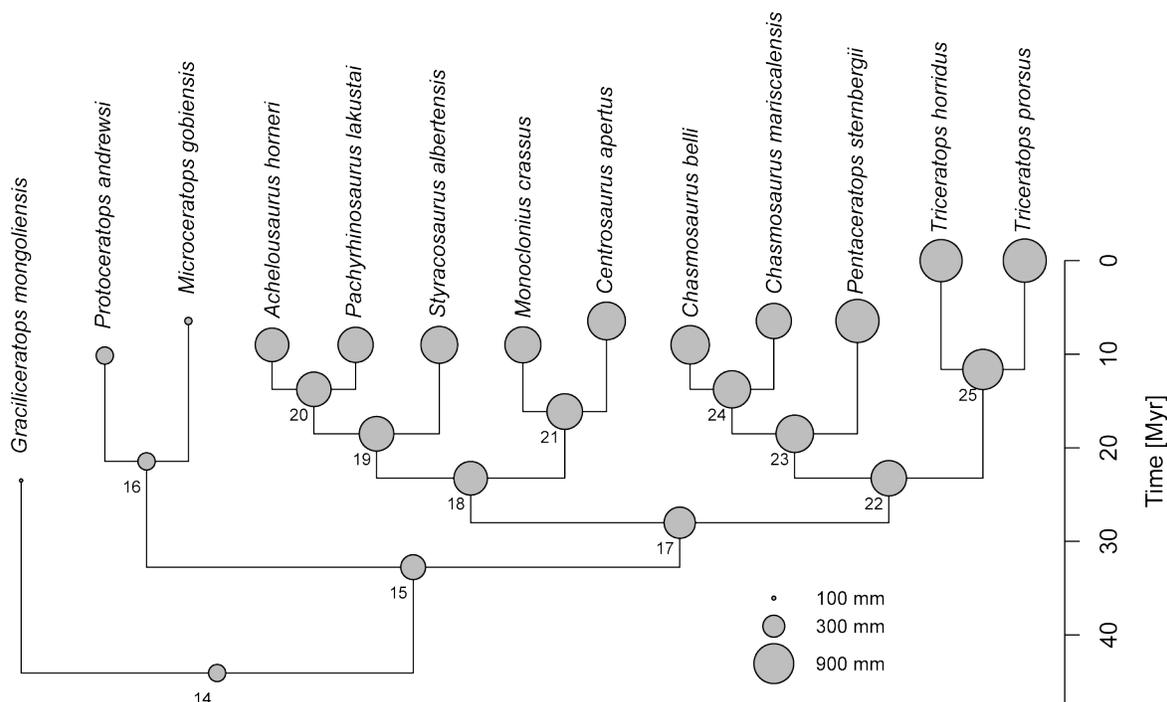


FIGURE 10.—Phylogeny of derived ceratopsians, with symbol sizes of terminal taxa that scale with log femur length. Symbol sizes for internal nodes are scaled the same way, using the maximum-likelihood estimates of ancestral trait values; note legend on the lower left side of the plot. Time axis on the right in million of years, counting backwards from the youngest terminal taxa. Internal nodes are labeled by number.

continually pulls wandering populations centripetally toward the long-term mean.

There are microevolutionary and macroevolutionary interpretations of OU models. The microevolutionary interpretation follows from the dynamics of populations on an adaptive landscape. With a few assumptions about the nature of variation and the topography of the adaptive landscape, the OU model describes the evolution of a population influenced by a stationary adaptive optimum (Lande, 1976; Arnold et al., 2001). Under this interpretation, gradients in natural selection push populations toward adaptive peaks and keep them in the vicinity of these peaks despite genetic drift, which would otherwise cause the populations to wander phenotypic space as a random walk. In rare circumstances, it is possible to apply this interpretation to fossil sequences and use it to assess and measure natural selection in an ancient population (Hunt et al., 2008). However, even very weak natural selection will cause a population to converge to an adaptive optimum in less than a few thousand generations, and most paleontological datasets cannot

resolve time this finely. For populations already near the adaptive optima, their dynamics reduce to the stasis model as described above when sampled at typical paleontological resolutions.

The macroevolutionary interpretation of the OU model will apply more often, especially in studies of entire clades. This formulation originated with Hansen (1997), who likened its dynamics to Simpson's notion of *adaptive zones* (Simpson, 1944). Here, the optimum is the mean phenotypic value within an adaptive zone. Species influenced by the same optimum differ somewhat according to species-specific adaptive adjustments, but do not wander far because the shared adaptive strategy of species in the same adaptive zone poses functional limits to divergence. When the attraction to the optimum is very strong, this model converges to the clade-level stasis model as described above. Clades can be assumed to share the same adaptive zone, or there may be shifts to different optima over time (Hansen, 1997), shifts on specific branches (Butler and King, 2004), or continual tracking of environmental or biotic variables (Hansen et al., 2008). Macroevolutionary OU

models are a particularly promising because they may be better able to capture some aspects of phenotypic evolution in real lineages, including the bounded nature of divergence within adaptive zones and morphological shifts associated with innovations.

OTHER ISSUES

Phylogenetic and Temporal Uncertainty.—We have thus far treated phylogeny as if it were a known quantity. In reality, phylogenies must be estimated from morphological, genetic, and possibly stratigraphic information (see Wagner and Marcot, this volume), and this estimation usually incurs considerable uncertainty. Phylogenetic analysis almost always produces a large set of trees of nearly equivalent support according to parsimony, likelihood, or Bayesian criteria. Although consensus trees are useful for summarizing the agreement among a set of trees, the polytomies they produce pose problems for analysis. Phylogenetic uncertainty (a “soft” polytomy) is not the same as the multifurcation that results when one species gives rise to two or more descendants (a “hard” polytomy). Treating soft polytomies as if they were hard assumes a set of relationships that are implied by none of the best-supported trees. A better strategy would be to repeat whatever analysis is desired on a random subsample of highly supported trees (e.g., equally parsimonious trees, or those within a certain range of tree lengths, log-likelihoods, or posterior probabilities). With true multifurcations, polytomies are an accurate reflection of relationships and should be retained through analysis.

The analogous uncertainty in ancestor–descendant sequences is age-model error. Unless it is determined by correlation across localities, the order of samples is usually clearly defined by stratigraphic superposition. As a result, the major source of error is in estimating the duration of time between each sample. Simulations can explore the robustness of results to varying degrees of inaccuracy or noise in the age models (Hunt, 2008b).

Temporal Scaling in Phylogeny.—For the dinosaur analyses, we scaled branch lengths to geological time, in units of millions of years. The Brownian motion and trend models hold that divergence accrues smoothly with elapsed time. This is an assumption, not a necessity. If punctuated equilibrium is fundamentally

correct, however, evolutionary changes do not accumulate gradually but instead accrue only at speciation events. One way to model this notion is to set all branch lengths to be equal (the exact value does not matter, but lengths are usually set to one). This tactic yields the same expected change per splitting event, rather than the same expected change per million years. It is important to note, however, that setting all branches equal corresponds strictly to a punctuated model only when all species that existed in a clade are sampled. With missing taxa, observed branches span unsampled speciation events, and we no longer expect the same amount of change on each branch (Bokma, 2008). If speciation rate does not vary systematically through the tree, we expect the number of missing speciation events to be proportional to time, and this leads us back to scaling branch lengths by their geological durations.

One can envision a continuum between purely punctuational and purely gradual modes of change. Pagel (1998; 1999; 2002) uses a temporal scaling parameter (κ) for branch lengths that measures the position along this continuum. κ ranges from zero to one, and represents the power to which all branch lengths are raised. When $\kappa = 1$, branch lengths are unchanged from the original units; when $\kappa = 0$, all branch lengths become equal to one. Intermediate values of κ produce patterns intermediate between punctuated and gradual divergence. Pagel (2002) discusses two other parameters that modify branch lengths. One of these, λ , quantifies the degree of phylogenetic signal in a trait. When $\lambda = 1$, traits have high phylogenetic signal, as much as Brownian motion. When $\lambda = 0$, close relatives do not resemble each other, and this model becomes equivalent to the clade-level stasis (white noise) as described above. Pagel’s third scaling parameter, δ , can be used to model accelerating or decelerating rates of evolutionary change.

CONCLUSIONS

The two fundamental pillars of the analyses presented in this chapter are genealogical relationships among populations and models of phenotypic change. Together these determine the probabilities of different evolutionary outcomes. All methods that go beyond description must make some assumptions about underlying nature of reality; the advantage of explicit statistical models is that these assumptions are

stated clearly. And, when important assumptions are violated, new models can be sought that better reflect the observations to be explained. As in other branches of science, progress will likely result from repeated iterations of data acquisition and attempts to understand those data in terms of simple but informative models.

REFERENCES

- AKAIKE, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716-723.
- ANDERSON, D. R., K. P. BURNHAM, AND W. L. THOMPSON. 2000. Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management*, 64(4):912-923.
- ARBOUR, V. M., M. E. BURNS, AND R. L. SISSONS. 2009. A re-description of the ankylosaurid dinosaur *Dyoplosaurus acutosquameus* Parks, 1924 (Ornithischia: Ankylosauria) and a revision of the genus. *Journal of Vertebrate Paleontology*, 29(4):1117-1135.
- ARNOLD, S. J., M. E. PFRENDER, AND A. G. JONES. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica*, 112-113:9-32.
- AVERIANOV, A. O., A. V. VORONKEVICH, S. V. LECHCHINSKIY, AND A. FAYNGERTZ. 2006. A ceratopsian dinosaur *Psittacosaurus sibiricus* from the Early Cretaceous of west Siberia, Russia and its phylogenetic relationships. *Journal of Systematic Palaeontology*, 4(4):359-395.
- BLOMBERG, S. P., T. GARLAND, AND A. R. IVES. 2003. Testing for phylogenetic signal in comparative data: behavioural traits are more labile. *Evolution*, 57(4):717-745.
- BOKMA, F. 2008. Detection of "Punctuated Equilibrium" by Bayesian Estimation of Speciation and Extinction Rates, Ancestral Character States, and Rates of Anagenetic and Cladogenetic Evolution on a Molecular Phylogeny. *Evolution*, 62(11):2718-2726.
- BOOKSTEIN, F. L. 1987. Random walk and the existence of evolutionary rates. *Paleobiology*, 13(4):446-464.
- BOYD, C. A., C. M. BROWN, R. D. SCHEETZ, AND J. A. CLARKE. 2009. Taxonomic revision of the basal neornithischian taxa *Thescelosaurus* and *Bugenasaura*. *Journal of Vertebrate Paleontology*, 29(3):758-770.
- BRUSATTE, S. L., M. J. BENTON, M. RUTA, AND G. T. LLOYD. 2008. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. *Science*, 321(5895):1485-1488.
- BURNHAM, K. P., AND D. R. ANDERSON. 2010. *Model Selection and Multimodel Inference*. Springer, New York, 516 p.
- BUTLER, M. A., AND A. A. KING. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist*, 164(6):683-695.
- BUTLER, R. J., P. M. GALTON, L. B. PORRO, L. M. CHIAPPE, D. M. HENDERSON, AND G. M. ERICKSON. 2010. Lower limits of ornithischian dinosaur body size inferred from a new Upper Jurassic heterodontosaurid from North America. *Proceedings of the Royal Society of London B*, 277:375-381.
- CARPENTER, K., J. BARTLETT, J. BIRD, AND R. BARRICK. 2008. Ankylosaurs from the Price River Quarries, Cedar Mountain Formation (Lower Cretaceous), east-central Utah. *Journal of Vertebrate Paleontology*, 28(4):1089-1101.
- CARRANO, M. T. 2006. Body-size evolution in the Dinosauria, p. 225-268. *In* M. T. Carrano, T. J. Gaudin, R. W. Blob, and J. R. Wible (eds.), *Amniote Paleobiology*. University of Chicago Press, Chicago.
- CLEVELAND, W. S. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829-836.
- CLYDE, W. C., AND P. D. GINGERICH. 1994. Rates of evolution in the dentition of early Eocene *Cantius*: comparison of size and shape. *Paleobiology*, 20(4):506-522.
- DALLA VECCHIA, F. M. 2009. *Tethyshadros insularis*, a new hadrosauroid dinosaur (Ornithischia) from the Upper Cretaceous of Italy. *Journal of Vertebrate Paleontology*, 29(4):1100-1116.
- DINIZ-FILHO, J. A. F., L. M. BINI, AND B. A. HAWKINS. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, 12:53-64.
- DRUMMOND, A., S. HO, M. PHILLIPS, AND A. RAMBAUT. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4:e88.
- ELDRIDGE, N., AND S. J. GOULD. 1972. Punctuated equilibria: an alternative to phyletic gradualism, p. 82-115. *In* T. J. M. Schopf (ed.), *Models in Paleobiology*. Freeman, Cooper & Company, San Francisco.
- ELDRIDGE, N., J. N. THOMPSON, P. M. BRAKEFIELD, S. GAVRILETS, D. JABLONSKI, J. B. C. JACKSON, R. E. LENSKI, B. S. LIEBERMAN, M. A. McPEEK, AND W. I. MILLER. 2005. The dynamics of evolutionary stasis. *Paleobiology*, 31(Supplement 2):133-145.
- FELSENSTEIN, J. 1985. Phylogenies and the comparative method. *American Naturalist*, 125(1):1-15.
- FELSENSTEIN, J. 2004. *Inferring Phylogenies*. Sinauer Associates, Inc., Sunderland, Massachusetts, 664 p.
- 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. *Systematic Biology*, 55(2):301-313.

- FISHER, D. C. 1994. Stratocladistics: Morphological and temporal patterns and their relation to phylogenetic process, p. 133-171. *In* L. Grande and O. Rieppel (eds.), *Interpreting the Hierarchy of Nature*. Academic Press.
- FOOTE, M. 1991. Analysis of morphological data, p. 59-86. *In* N. L. Gilinsky and P. W. Signor (eds.), *Analytical Paleobiology*. Volume 4. The Paleontological Society.
- FOOTE, M. 1996. On the probability of ancestors in the fossil record. *Paleobiology*, 22(2):141-151.
- FORD, T. L., AND J. I. KIRKLAND. 2001. Carlsbad ankylosaur (*Ornithischia*, *Ankylosauria*): an ankylosaurid and not a nodosaurid, p. 239-260. *In* K. Carpenter (ed.), *The Armored Dinosaurs*. Indiana University Press, Bloomington, IN.
- GARLAND, T., P. H. HARVEY, AND A. R. IVES. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology*, 41(1):18-32.
- GARLAND, T., AND A. R. IVES. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *American Naturalist*, 155(3):346-364.
- GEARY, D. H., G. HUNT, I. MAGYAR, AND H. SCHULTZ. 2010. The paradox of gradualism: phyletic evolution in two lineages of *lymnocardiid* bivalves (Lake Pannon, central Europe). *Paleobiology*. In press.
- GINGERICH, P. D. 1983. Rates of evolution: effects of time and temporal scaling. *Science*, 222(4620):159-161.
- GINGERICH, P. D. 1993. Quantification and comparison of evolutionary rates. *American Journal of Science*, 293-A:453-478.
- GINGERICH, P. D. 2000. Arithmetic or geometric normality of biological variation: an empirical test of theory. *Journal of Theoretical Biology*, 204(2):201-221.
- GINGERICH, P. D. 2009. Rates of evolution. *Annual Review of Ecology Evolution and Systematics*, 40:657-675.
- GOULD, S. J. 2002. *The Structure of Evolutionary Theory*. Belknap Press, Cambridge, Massachusetts, 1433 p.
- HANNISDAL, B. 2006. Phenotypic evolution in the fossil record: Numerical experiments. *Journal of Geology*, 114(2):133-153.
- HANSEN, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51(5):1341-1351.
- HANSEN, T. F., AND E. P. MARTINS. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, 50(4):1404-1417.
- HANSEN, T. F., J. PIENAAR, AND S. H. ORZACK. 2008. A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, 62(8):1965-1977.
- HARMON, L. J., J. WEIR, C. BROCK, R. GLOR, W. CHALLENGER, AND G. HUNT. 2009. *geiger*: analysis of evolutionary diversification.
- HONE, D. W. E., T. M. KEESEY, D. PISANI, AND A. PURVIS. 2005. Macroevolutionary trends in the Dinosauria: Cope's Rule. *Journal of Evolutionary Biology*, 18:587-595.
- HUNT, G. 2006. Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology*, 32(4):578-601.
- HUNT, G. 2008a. Evolutionary patterns within fossil lineages: model-based assessment of modes, rates, punctuations and process, p. 117-131. *In* R. K. Bambach and P. H. Kelley (eds.), *From Evolution to Geobiology: Research Questions Driving Paleontology at the Start of a New Century*. The Paleontological Society.
- HUNT, G. 2008b. Gradual or pulsed evolution: when should punctuational explanations be preferred? *Paleobiology*, 34(3):360-377.
- HUNT, G. 2008c. *paleoTS*: Modeling evolution in paleontological time-series.
- HUNT, G., M. A. BELL, AND M. P. TRAVIS. 2008. Evolution toward a new adaptive optimum: Phenotypic evolution in a fossil stickleback lineage. *Evolution*, 62(3):700-710.
- HUNT, G., T. M. CRONIN, AND K. ROY. 2005. Species-energy relationship in the deep sea: a test using the Quaternary fossil record. *Ecology Letters*, 8:739-747.
- LANDE, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30:314-334.
- LYNCH, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *American Naturalist*, 136(6):727-741.
- MADDISON, W. P. 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogeny. *Systematic Zoology*, 40:304-314.
- MADDISON, W. P., AND D. R. MADDISON. 2009. *Mesquite*: a modular system for evolutionary analysis.
- MAIDMENT, S. C. R., D. B. NORMAN, P. M. BARRETT, AND P. UPCHURCH. 2008. Systematics and phylogeny of *Stegosauria* (Dinosauria: *Ornithischia*). *Journal of Systematic Palaeontology*, 6(04):367-407.
- MARCOT, J. D., AND D. L. FOX. 2008. *StrataPhy*: a new computer program for stratocladistic analysis. *Palaeontologia Electronica*, 11(1):1-16.
- MARTINS, E. P., AND T. F. HANSEN. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist*, 149(4):646-667.
- McKINNEY, M. L. 1990. Classifying and analyzing evolutionary trends, p. 28-58. *In* K. J. McNamara (ed.), *Evolutionary Trends*. University of Arizona Press, Tucson.
- McSHEA, D. W. 1994. Mechanisms of large-scale evolutionary trends. *Evolution*, 48(6):1747-1763.

- NOVAS, F. E., A. V. CAMBIASO, AND A. AMBROSIO. 2004. A new basal iguanodontian (Dinosauria, Ornithischia) from the Upper Cretaceous of Patagonia. *Ameghiniana*, 41(1):75-82.
- O'MEARA, B. C., C. ANÉ, M. J. SANDERSON, AND P. C. WAINWRIGHT. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution*, 60(5):922-933.
- PAGEL, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 255(1342):37-45.
- PAGEL, M. 1998. Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, 26:331-348.
- PAGEL, M. 1999. Inferring the historical patterns of biological evolution. *Nature*, 401:877-884.
- PAGEL, M. 2002. Modelling the evolution of continuously varying characters on phylogenetic trees: the case of hominid cranial capacity, p. 269-286. *In* N. MacLeod and P. L. Forey (eds.), *Morphology, Shape and Phylogeny*. Taylor & Francis, London.
- PARADIS, E., AND J. CLAUDE. 2002. Analysis of comparative data using generalized estimating equations. *Journal of Theoretical Biology*, 218:175-185.
- PARADIS, E., J. CLAUDE, AND K. STRIMMER. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20(2):289-290.
- R DEVELOPMENT CORE TEAM. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- RAUP, D. M. 1977. Stochastic models in evolutionary paleobiology, p. 59-78. *In* A. Hallam (ed.), *Patterns of Evolution as Illustrated by the Fossil Record*. Elsevier Scientific Publishing Company, Amsterdam.
- RAUP, D. M., AND R. E. CRICK. 1981. Evolution of single characters in the Jurassic ammonite *Kosmoceras*. *Paleobiology*, 7(2):200-215.
- RAUP, D. M., AND S. J. GOULD. 1974. Stochastic simulation and evolution of morphology: towards a nomothetic paleontology. *Systematic Zoology*, 23:305-322.
- REVELL, L. J., AND D. C. COLLAR. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution*, 63(4):1090-1100.
- REVELL, L. J., AND L. J. HARMON. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evolutionary Ecology Research*, 10(3):311-331.
- ROHLF, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution*, 55(11):2143-2160.
- ROOPNARINE, P. D. 2001. The description and classification of evolutionary mode: a computational approach. *Paleobiology*, 27(3):446-465.
- ROOPNARINE, P. D., G. BYARS, AND P. FITZGERALD. 1999. Anagenetic evolution, stratophenetic patterns, and random walk models. *Paleobiology*, 25(1):41-57.
- ROY, K., AND M. FOOTE. 1997. Morphological approaches to measuring biodiversity. *Trends in Ecology and Evolution*, 12(7):277-281.
- 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 B-Biological Sciences*, 273(1598):2107-2111.
- RYAN, M. J. 2007. A new basal centrosaurine ceratopsid from the Oldman Formation, southeastern Alberta. *Journal of Paleontology*, 81(2):376-396.
- SANDERSON, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, 19(2):301-302.
- SCHLUTER, D., T. PRICE, A. MOOERS, AND D. LUDWIG. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution*, 51(6):1699-1711.
- SHEETS, H. D., AND C. E. MITCHELL. 2001. Why the null matters: statistical tests, random walks and evolution. *Genetica*, 112-113:105-125.
- SIDLAKUSKAS, B. 2008. Continuous and Arrested Morphological Diversification in Sister Clades of Characiform Fishes: A Phylomorphospace Approach. *Evolution*, 62(12):3135-3156.
- SIMPSON, G. G. 1944. *Tempo and Mode in Evolution*. Columbia University Press, New York, 237 p.
- SMITH, A. 1994. *Systematics and the Fossil Record*. Blackwell Scientific, Oxford.
- SUES, H.-D., AND A. AVERIANOV. 2009. A new basal hadrosauroid dinosaur from the Late Cretaceous of Uzbekistan and the early radiation of duck-billed dinosaurs. *Proceedings of the Royal Society B: Biological Sciences*, 276(1667):2549-2555.
- THOMAS, G. H., S. MEIRI, AND A. B. PHILLIMORE. 2009. Body size diversification in *Anolis*: novel environment and island effects. *Evolution*, 63(8):2017-2030.
- VICKARYOUS, M. K., T. MARYANSKA, AND D. B. WEISHAMPEL. 2004. Ankylosauria, p. 363-392. *In* D. B. Weishampel, P. Dodson, and H. Osmólska (eds.), *The Dinosauria*, second edition. University of California Press, Berkeley.
- WALKER, J. D., AND J. W. GEISSMAN. 2009. 2009 GSA geological time scale. *GSA Today*, 19(4/5):60-61.

- WEBSTER, A. J., AND A. PURVIS. 2002a. Ancestral states and evolutionary rates of continuous characters, p. 247-268. *In* N. MacLeod and P. L. Forey (eds.), *Morphology, Shape and Phylogeny*. Taylor & Francis, London.
- WEBSTER, A. J., AND A. PURVIS. 2002b. Testing the accuracy of methods for reconstructing ancestral states of continuous characters. *Proceedings of the Royal Society of London, Series B*, 269:143-149.
- WEISHAMPEL, D. B., C.-M. JIANU, Z. CSIKI, AND D. B. NORMAN. 2003. Osteology and phylogeny of *Zalmoxes* (n. g.), an unusual euornithopod dinosaur from the latest Cretaceous of Romania. *Journal of Systematic Palaeontology*, 1(2):65-123.
- YOU, H.-L., K. TANOUE, AND P. DODSON. 2008. New data on cranial anatomy of the ceratopsian dinosaur *Psittacosaurus major*. *Acta Palaeontologica Polonica*, 53(2):183-196.

